

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 17, 2003, 12:13:06 ; Search time 10.3726 Seconds
(without alignments)
3378.605 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263
Perfect score: 1386
Sequence: 1 MNSYPSKLPIDKSHSLQLQ.....NKNNEVNSEPALTDMLKR 263

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A:Geneseq_101002:*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386	100.0	1664	20	AAW99462
2	1386	100.0	1664	23	AAU79331
3	1386	100.0	1664	23	AAE19799
4	241	17.4	236	20	AAE19456
5	241	17.4	236	23	AAE19800
6	138.5	10.0	1166	20	AAE19800
7	133	9.6	767	22	AAE19800
8	133	9.6	971	19	AAW48896
9	131.5	9.5	1174	22	AAE19800
10	131.5	9.5	2439	22	AAE19800

11	127.5	9.2	1335	23	ABP38871
12	127.5	9.2	1542	22	ABP71456
13	127	9.2	4134	20	AAV31946
14	126	9.1	1817	21	AAE42230
15	125	9.0	722	22	ABP63899
16	125	9.0	1545	22	ABP65577
17	124.5	9.0	811	22	ABP62661
18	124	8.9	1798	22	ABP71695
19	124	8.9	2781	21	AAV57453
20	123	8.9	6772	22	AAV93543
21	123	8.9	874	22	ABP62601
22	123	8.9	1557	22	ABP63735
23	123	8.9	4019	22	AAE13839
24	122.5	8.8	2703	22	ABP60074
25	122	8.8	761	22	ABP67519
26	120.5	8.7	1069	22	ABP61305
27	120.5	8.7	1428	22	ABP70377
28	120	8.7	844	22	AAU35904
29	120	8.7	849	18	AAW20655
30	120	8.7	1744	22	ABP62890
31	120	8.7	1908	22	ABP70137
32	120	8.7	1954	22	ABP59197
33	119.5	8.6	690	22	ABP71683
34	119.5	8.6	1888	22	AAU31049
35	119.5	8.6	1920	22	ABP64441
36	119	8.6	1186	22	ABP64079
37	118.5	8.5	696	22	ABP60920
38	118	8.5	738	22	ABP58060
39	118	8.5	2907	21	AAV57452
40	118	8.5	10182	23	ABP38314
41	117.5	8.5	3502	22	ABP58382
42	117	8.4	5464	22	ABG11810
43	117	8.4	7718	22	ABG11811
44	116.5	8.4	1379	22	ABP68840
45	116	8.4	797	22	ABP62042

ALIGNMENTS

RESULT 1	
AAW99462	AAW99462 standard; Protein: 1664 AA.
XX	XX
AC	AAW99462;
DT	08-JUN-1999 (first entry)
DE	C.albicans alpha-INTip protein.
XX	XX
KW	Integrin-like motif; vaccine; immune response; antibody; inhibition;
KW	adhesion; endothelial cell; pathogenesis; infection; probe.
OS	Candida albicans.
XX	XX
PN	US5886151-A.
PD	23-MAR-1999.
XX	XX
PF	03-MAY-1996; 96US-0642846.
XX	XX
PR	03-MAY-1996; 96US-0642846.
XX	XX
PA	(MIND) UNIV MINNESOTA.
PI	Bendel CM, Gale CA, Hostetter MK, Kendrick K, Tao NJ;
DR	WPI; 1999-242618/20.
XX	XX
DR	N-PSDB; AAX25885.
XX	XX
PT	New isolated Candida albicans protein with Integrin-like motifs
XX	XX
PS	Examples: Column 13-14; 21pp; English.

Staphylococcus epi
Drosophila melanog
Plasmodium falcipa
Human ORF1994
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Human transcriptio
Human polypeptide,
Drosophila melanog
Drosophila melanog
Human lung tumour-
Drosophila melanog
Drosophila melanog
Drosophila melanog
Helicobacter pylori
H. pylori cytoplas
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Novel human secre
Drosophila melanog
Drosophila melanog
Drosophila melanog
Human transcriptio
Staphylococcus epi
Drosophila melanog
Novel human diagno
Drosophila melanog
Drosophila melanog


```

FT Misc-difference 251
FT /note= "Encoded by CTG"
FT Misc-difference 295
FT /note= "Encoded by CTG"
FT Misc-difference 339
FT /note= "Encoded by CTG"
FT Misc-difference 519
FT /note= "Encoded by CTG"
FT Misc-difference 1102
FT /note= "Encoded by CTG"
FT Region
FT 1149..1151
FT /note= "RGD site"
PN US6346411-B1.
PD 12-FEB-2002.
PF 08-MAR-1999; 99US-0264604.
PR 03-MAY-1996; 96US-0642846.
XX (MINU ) UNIV MINNESOTA.
XX Hostetter MK, Gale CA, Bendel CM, Tao N;
XX WPI: 2002-224995/28.
XX N-PSDB: AAD31519.
XX Polynucleotide encoding Candida albicans protein with integrin-like
XX motifs and protein, useful as vaccines and for raising antibodies for
XX inhibiting adhesion of pathogen to cells preferably epithelial cells
XX
PS Claim 3; Column 15-18; 25pp; English.
CC The present invention relates to an isolated polynucleotide encoding
CC Candida albicans protein with integrin-like motifs. Candida albicans
CC protein is utilised in vaccines (for gastrointestinal pathogens like
CC cholera) and as antigens to prepare anti-peptide antibodies, which are
CC utilised in inhibiting adhesion of C. albicans to cells preferably
CC epithelial cells. In the genitourinary tract, expression of spermicides
CC by S. cerevisiae transformed with the C. albicans integrin-like gene on
CC an extrachromosomal plasmid could provide a cheap and infrequent method
CC of contraception. Also, synthesis of protein-based antiretroviral
CC agents could help to reduce transmission of human immunodeficiency virus
CC (HIV) in the birth canal. The present sequence is Candida albicans
CC integrin-like protein (alphaIntlp).
CC
XX
XX Sequence 1664 AA;
SQ
Query Match 100.0%; Score 1386; DB 23; Length 1664;
Best Local Similarity 100.0%; Pred. No. 5e-107;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNSTPSKLLPDKHSHLQLOPOSSASIFNSPTKPLNFRNTSKPSLDNSSSDTYTSEQ 60
DB 1 MNSTPSKLLPDKHSHLQLOPOSSASIFNSPTKPLNFRNTSKPSLDNSSSDTYTSEQ 60
OY 61 DOEKKEEKKDTAFQTSFPRNFDLNSIDIOQTIGHQOQPOQOQOLSTDNKLDEESF 120
DB 61 DOEKKEEKKDTAFQTSFPRNFDLNSIDIOQTIGHQOQPOQOQOLSTDNKLDEESF 120
OY 121 CTPMTSTDLTKQNTPTVDKVNENHAPTYINTSPNKSIMKATPKASPKVAFVTNPEIH 180
DB 121 CTPMTSTDLTKQNTPTVDKVNENHAPTYINTSPNKSIMKATPKASPKVAFVTNPEIH 180
OY 181 HYPDNRVEEDDSQCKEDSVEPPLIGHQWKDPSQFNYSDEDTNASVPTPPLHTTKPTFA 240
DB 181 HYPDNRVEEDDSQCKEDSVEPPLIGHQWKDPSQFNYSDEDTNASVPTPPLHTTKPTFA 240
OY 241 QLLNKNNEVNSPEALTDMLKLR 263
DB 241 QLLNKNNEVNSPEALTDMLKLR 263

```

```

RESULT 4
AAW9456
ID AAW9456 standard; protein; 236 AA.
XX
XX AAW9456;
AC AAW9456;
XX
XX 08-JUN-1999 (first entry)
DT
XX
XX Amino acids 218-453 of C.albicans integrin-like protein 1.
DE
XX
XX Integrin-like motif; vaccine; immune response; antibody; inhibition;
XX adhesion; endothelial cell; pathogenesis; infection.
XX
XX Candida albicans.
OS
XX
XX US5886151-A.
PN
XX
XX 23-MAR-1999.
PD
XX
XX 03-MAY-1996; 96US-0642846.
PF
XX
XX 03-MAY-1996; 96US-0642846.
PR
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Bendel CM, Gale CA, Hostetter MK, Kendrick K, Tao N;
XX WPI: 1999-242618/20.
XX
XX New isolated Candida albicans protein with integrin-like motifs
XX
XX
XX Claim 1; Column 35; 21pp; English.
XX
XX Peptides AAW9456-w99461 are derived from a Candida albicans protein
XX with integrin-like motifs, alpha-INP1. This sequence represents amino
XX acids 218-453 of alpha-INP1. The peptides can be used for producing
XX vaccines for stimulating an immune response. The antibodies can inhibit
XX the adhesion of C.albicans to cells, particularly endothelial cells.
XX This blocking activity of the adhesion to cells can reduce or prevent
XX subsequent events in the pathogenesis of invasive candidal infection.
XX
XX
XX Sequence 236 AA;
SQ
Query Match 17.4%; Score 241; DB 20; Length 236;
Best Local Similarity 100.0%; Pred. No. 2.8e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 1; Gaps 0;
OY 218 SDEDTNASVPTPPLHTTKPTFAQLLNKNNEVNSPEALTDMLKLR 263
DB 1 SDEDTNASVPTPPLHTTKPTFAQLLNKNNEVNSPEALTDMLKLR 46
RESULT 5
AAE19800
ID AAE19800 standard; protein; 236 AA.
XX
XX AAE19800;
AC AAE19800;
XX
XX 18-JUN-2002 (first entry)
DT
XX
XX Candida albicans integrin-like protein (alphaIntlp) fragment.
DE
XX
XX Integrin-like protein; alphaIntlp; gene; fungicide; contraceptive;
XX HIV; human immunodeficiency virus; vaccine; cholera.
XX
XX Candida albicans.
OS
XX
XX US6346411-B1.
PN
XX
XX 12-FEB-2002.
PD
XX
XX 08-MAR-1999; 99US-0264604.
XX

```


XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX MO200171042-A2.
XX PD 27-SEP-2001.
XX PD 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX N-PSDB: ABL02343.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure: SEQ ID NO 1512; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutic and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 767 AA:
XX
XX Query Match 9.6%; Score 133; DB 22; Length 767;
XX Best Local Similarity 24.1%; Pred. No. 0.014;
XX Matches 87; Conservative 31; Mismatches 121; Indels 122; Gaps 16;
XX
XX 2 NSTPSKLPIDKHSHLQLOPOSSASIFNS-----PTKPLNPRTNSKP-----SLDPNS 51
XX 232 NSQPSVLNLAQLNALIGDASAAATPVGSSNTSSNPNPPIORHSQSLGSLSPSS 291
XX 52 SSDTYTSRDODEKGR-----EKKRTAFOTSFDRNFDLDSIDIOQ----- 92
XX 292 ISNMSSASCHGSGSLSGNNGNNGSTTNTAILTS-----NGNVLDQLSAGNVV 343
XX
XX 93 -----TIOHQOQOP-----QOQOOLSO-----TDNNLIDE-----FSQOTPMSTYL--- 128
XX DB 344 TTINLPTQHQHQOQOPQOQOQOQOQHONTATGNSNNNNLMRSTLGLVSLKTNNTPRAKSV 403
XX QY 139 -DLTKONPVDKV-----NENH-----APTYINTS---PKSINK 159
XX DB 404 SLMAMANTASKILKATLVANGSSMDIVQOQOQHNTTTTGTQVTLNGSKPLPNLVNNG 463
XX QY 160 KATPKASPK-----VAFVTYNPEIHNY-----PDNVEEEDSOQKE 197
XX DB 464 ILTPRTSPKNTOTLFTTTNGSTOQOQALVASLSKTNILLQEPYQLVQOQOQOQOQO 523
XX QY 198 DSEVPEPL-----IOHQMKPSOFNSDEDTNASVPTPLHTTKPTFAQLLNKNNEVNSE 252
XX DB 524 QOQOQOQOQOQOQOQHNVLPISPTSSVSGSSKSSSPTP-----QOQOQOQLLKQKLELIE 578
XX QY 253 P 253
XX DB 579 P 579

RESULT 8
AAW48896
ID AAW48896 standard; Protein; 971 AA.
XX
XX AAW48896;
XX AC
XX 13-OCT-1998 (first entry)
XX DT
XX
XX Candida albicans CaCLA4 protein.
XX DE
XX
XX CaCLA4; protein kinase; Ste20p family; screening; virulence;
XX hyphal formation; pathogenic fungi; inhibitor; inflammation;
XX antilymphotic.
XX KW
XX
XX Candida albicans.
XX OS
XX
XX WO9818927-A1.
XX PN
XX 07-MAY-1998.
XX PD
XX
XX 29-OCT-1997; 97WO-CA00809.
XX PF
XX 30-OCT-1996; 96US-0029458.
XX PR
XX
XX (CANADA) NAT RES COUNCIL CANADA.
XX PA
XX
XX Leberer E, Thomas DY;
XX PI
XX WPI: 1998-272222/24.
XX DR
XX N-PSDB: AAW32554.
XX
XX In vitro screening test for agents that inhibit Candida genes
XX involved in virulence - and transition to hyphal form, potentially
XX useful as antilymphotic agents
XX PT
XX
XX Disclosure: Fig 7; 79pp; English.
XX
XX The sequence is that of the CaCLA4 protein which can be used
XX in the development of an in vitro screening test for compounds
XX that inhibit biological activity of the protein and a system for
XX measuring its activity. The protein is involved in virulence and
XX hyphal formation. Inhibitors are potentially useful for rendering
XX pathogenic fungi (any species in which hyphal induction by kinase
XX occurs) avirulent and/or to treat inflammation.
XX CC
XX
XX Sequence 971 AA:
XX
XX Query Match 9.6%; Score 133; DB 19; Length 971;
XX Best Local Similarity 24.3%; Pred. No. 0.019;
XX Matches 63; Conservative 33; Mismatches 85; Indels 78; Gaps 12;
XX
XX 4 TFSKLPIDKHSHLQLOPOSSASIFNSPTK-----PLN-----FPR 40
XX DB 358 TPEHLTLQNGSSH-----QHTSSSGSLSPSGNNNNNNSTNNNTKYNVPLNNLMKSELIA 414
XX QY 41 TNSKESLDPNSSSDTYTSRDODEKGRKKRTAFOTSFDRNFDLDSIDIOQTIOHQOQO 100
XX DB 415 RRAPEPTSGTSDTYSNKNHODRSGYEDQ-----RQGRDSSOQOQOQOQOQOQO 465
XX QY 101 PQOQOOLQOTDNNLIDEFSPOTPMT-----STLDLTKONPVDKVNENHAPTYINTSPNS 156
XX DB 466 KSQO 502
XX QY 157 IMKATPKASPKKVAFTVYNPEIHNPDNVEEEDSOQKEDESEVPEPLIOHQMKD-PSQF 215
XX DB 503 -TQAAAGKXWPSKI-----HPDLKIQGCTNNYIKSSCTDANQYDGAQKQKIFP 550
XX QY 216 NY-----SDEDTNASVPTPP 231
XX DB 551 NLQKKSOOOLASKOPSP 569

RESULT 9
AAG85039
ID AAG85039 standard; Protein: 1174 AA.
XX
AC AAG85039;
XX
DT 11-SEP-2001 (first entry)
XX
DE Shrimp white spot Bacilliform virus (WSBV) protein 130.
XX
KM Shrimp white spot Bacilliform virus; WSBV: diagnosis; viral infection;
KM antiviral agent; gene expression; antisense construct;
KM transgenic viral resistant shrimp.
XX
OS White spot syndrome virus.
XX
PN WO200138351-A2.

31-MAY-2001.

PF 08-NOV-2000; 2000WO-US28888.

PR 24-NOV-1999; 99CN-0124717.

XX
XX (PENY-) PE CORP NY.
XX (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX (SINO-) SINOGENOMAX CO LTD.

XX
XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX WPI; 2001-355877/37.
XX N-PSDB; AAG62819.

XX
XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
XX (WSBV), useful for producing viral polypeptides that can be used to
XX screen for agents that are useful for treating WSBV infection -

Claim 1; Figure 3; 626pp; English.

XX
XX The invention provides the primary nucleotide sequence of the WSBV genome
XX (AAG62819), predicted transcript sequences (AAG62819-AAG62819) and
XX encoded proteins (AAG62819-AAG62819) and oligonucleotide sequences
XX (AAG62819-63160) suitable for use as primers or probes. The nucleic acid
XX molecules and proteins of the invention are useful for diagnosis and
XX monitoring viral infection, in screens for antiviral agents and for
XX monitoring viral gene expression or activity during a treatment regimen.
XX The nucleic acid molecules are also useful as antisense constructs to
XX control viral gene expression in infected cells and tissues and to create
XX transgenic viral resistant shrimp.

SQ Sequence 1174 AA;

Query Match 9.5%; Score 131.5; DB 22; Length 1174;

Best Local Similarity 20.4%; Pred. No. 0.033;

Matches 58; Conservative 51; Mismatches 106; Indels 69; Gaps 11;

XX
XX 10 PIDKHSIQLQPOSSASIFNSPTKPLNFPRTNSKP-----SLDPNSSSD 54

XX
XX 240 PDEGKN---EPKDEDAIEN---ESVSHSHVESPSSEGNDGMDYFSSIAGGNDN 293

XX
XX 55 TYTSEDOQEKKEKKTAFOTSPDRNFDLNSIDIQOTIQHQOQPOQOQOLSOTDNNL 114

XX
XX 294 EDE 350

XX
XX 115 IDESFQPMSTLDITKONFTVDKVNENHAPTYINTSPKNSIM---KKATPKA----- 165

XX
XX 351 EDEISTHMOENNAKRRKSPRPDIQECEDAVVPPPIKKTIDILPOLKEPSKAPRPFESI 410

XX
XX 166 -----SPKKAFTVYNPEIHHYPNRVEEDOSQOKEDEV-----PP 203
XX 411 LGSGGEGYDQNLDIAPRPVSIYTFP-----PDNKEGEERDLMDOQSMMLPPPPPP 465

XX
XX 204 LIHQWKDPQSQFNYSDEDTNNAVPPPTPLHTTKPTFAQLKNKN 247
XX 466 PPPHQ---PQL---KPTNILLPPPPPPPTNQSLEFN---NNNN 500

RESULT 10

ABB68896

ID ABB68896 standard; Protein: 2439 AA.

XX ABB68896;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 33480.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL12999.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure; SEQ ID NO 33480; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (AAB57737-AB572072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 2439 AA;

Query Match 9.5%; Score 131.5; DB 22; Length 2439;

Best Local Similarity 19.9%; Pred. No. 0.085;

Matches 63; Conservative 38; Mismatches 109; Indels 107; Gaps 11;

XX
XX 10 PIDKHSIQLQ-----QPOSSASIFNSPTKPLNFPRT-----NSKPSL----- 47

XX
XX 1009 PVSKANHSNFSAGDNPSSRN---SSPLIGINSPOSVANAGAPSKKAPNLLRGKLO 1064

XX
XX 48 ----DPNSSDPTYTSEDOQEKKEKKTAFOTSPDRNFDLDN-----5----- 86

XX
XX 1065 OGVERKAKKDPAAKPKKREKAKGVKTKAADEKIKLLEKYETTPAAGAGYEGQIT 1124

XX
XX 87 -----SIDIQOTIQHQOQPOQOQOLSOTDNNLIDFSFOT 122

XX
XX 1125 NQOQOQOTIVNNHMQQLPMOTQLALPTQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1184

XX
XX 123 PWTSTL-----DLTKONFTVD-----KVENHNAFYIN-----TSPKSI 157

Db 1185 OOTSOLOOOPAOLOOQOOPQOQOQOQOQHHPVYNSFQOQOQOQOQOQOQOQVNPGE-- 1242
QY 158 MKKATPKSPKRYATFTVTPREIHHYBDNRVEEDSOQKEDSEVPELIOHMKD-PSQFN 216
Db 1243 -----PRVPELIDIRLGNK--HVVVKNRKDRKKGNQENASIDEROLKRSYSDPPQOL 1294
QY 217 YSEEDTNASVPPPTPLH 233
Db 1295 LMDPAESKOAKLTTPPH 1311

RESULT 11

ABP38871
ID ABP38871 standard; Protein; 1335 AA.

AC ABP38871;
XX

DT 24-JUL-2002 (first entry)
XX

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3716.
XX

KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
antibacterial; gene therapy.
XX

OS Staphylococcus epidermidis.
XX

PN US6380370-B1.
XX

PD 30-APR-2002.
XX

PF 13-AUG-1998; 98US-0134001.
XX

PR 14-AUG-1997; 97US-055779P.
08-NOV-1997; 97US-064964P.
XX

PA (GENO-) GENOME THERAPEUTICS CORP.
XX

PI Doucette-Stamm LA, Bush D;
XX

DR WPI: 2002-381255/41.
N-PSDB: ABN91416.
XX

PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
polypeptide, useful for diagnosing and treating bacterial infections -
XX

PS Disclosure: SEQ ID 3716; 267pp; English.
XX

CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
frame (ORF) nucleic acid sequences which encode the amino acid sequences
given in ABP3124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX

SQ Sequence 1335 AA;
XX

Query Match 9.2%; Score 127.5; DB 23; Length 1335;
Best Local Similarity 21.2%; Pred. No. 0.084; Indels 35; Gaps 7;

Matches 54; Conservative 41; Mismatches 125; Indels 35; Gaps 7;

QY 18 OLQPOSSASIFNSPTKPLNFPRTNSKPSIDPNSSSDTYT-SEDOEKGKEKKDAFQT 76
Db 32 QPQNSNHNKAVDDQFALQAEKASEVQSTNVSGTQYODPTQVQPKQDQSTTYDA 91

QY 77 SFDRNFDLNSLIDIOQTIOHQOQOPOQOQOOLSDTNLIDFESF-OTPMSTLIDLRKPN 135
Db 92 SLID-----EMSTYNEISSNQKQSLSTDANQNTNSVTKNQGETNDLQEDK 140

QY 136 TVQKVENHAPTYINTSPMKSIKKATPKASPKVAFTVTPREIHHYBDNRVEED-OSQ 194
Db 141 TSDTNQLODETOSVAKENEKDLGAMANNEDQDKM--TASQSENALETQTASNDNESQ 198
QY 195 OKEDSV-----EPPELIOHMKDPSQFNYSDEDTNASVPPPTPLHTTKPTFAQLINKN 247
Db 199 QKSQVTSQNETATPKVSNSTNAGYNFDYDEDDDDSDTHLEPISL-----N 246
QY 248 EVNSEPEALIDMKL 262
Db 247 NVNATSKQTTSYK 261

RESULT 12

ABB71456
ID ABB71456 standard; Protein; 1542 AA.

AC ABB71456;
XX

DT 26-MAR-2002 (first entry)
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 41160.
XX

KM Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX

OS Drosophila melanogaster.
XX

PN WO200171042-A2.
XX

PD 27-SEP-2001.
XX

PF 23-MAR-2001; 2001WO-US09231.
XX

PR 23-MAR-2000; 2000US-191637P.
11-JUL-2000; 2000US-0614150.
XX

PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX

DR WPI: 2001-656860/75.
N-PSDB: ABL15559.
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
XX

PS Disclosure: SEQ ID NO 41160; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 1542 AA;
XX

Query Match 9.2%; Score 127.5; DB 22; Length 1542;
Best Local Similarity 23.5%; Pred. No. 0.1;

Matches 65; Conservative 45; Mismatches 68; Indels 99; Gaps 19;

QY 2 NSTP-SKLPIDKHSHLOPOSSSAS--TF-NSPTKPLNF---PRTNSKPSIDPNSSSD 54
Db 306 HATPQAIFFPTEVTTTKMTTSTSAVPVFTGEPPTPPAFSPICGTGI-PNIDP---AD 361

This Page Blank (uspto)

Shears, Beverly

From: Devi, Sarvamangala
Sent: Friday, March 14, 2003 4:38 PM
To: Shears, Beverly
Subject: 09/964,858

Hello Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 1 and a fragment comprising amino acid residues 1-263 of SEQ ID NO: 1 in case 09/964,858?

Thanks.

S. DEVI, Ph.D.
AU 1645
CM1-7E15

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

This Page Blank (uspto)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:21:44 ; Search time 4.09445 Seconds

(without alignments)
1889.932 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263

Perfect score: 1386

Sequence: 1 MNSTPSKLPIDKSHLQLO.....NKNNEVSEPEALTMKLR 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386	100.0	1664	1 US-09-599-652-2	Sequence 2, Appli
2	1386	100.0	1664	2 US-08-642-846-2	Sequence 2, Appli
3	1386	100.0	1664	4 US-09-264-604-2	Sequence 2, Appli
4	241	17.4	236	1 US-09-599-652-3	Sequence 3, Appli
5	241	17.4	236	2 US-08-642-846-3	Sequence 3, Appli
6	241	17.4	236	4 US-09-264-604-3	Sequence 3, Appli
7	127.5	9.2	1335	4 US-09-134-001C-3716	Sequence 3716, Ap
8	118	8.5	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
9	114	8.2	1164	4 US-08-923-992A-10	Sequence 218, App
10	112.5	8.1	565	4 US-08-961-083-218	Sequence 2, Appli
11	112	8.1	984	1 US-08-242-932-2	Sequence 2, Appli
12	112	8.1	984	1 US-08-714-481-2	Sequence 2, Appli
13	112	8.1	984	5 PCT-US95-06111-2	Sequence 2, Appli
14	111	8.0	652	4 US-09-134-001C-3517	Sequence 3517, Ap
15	111	8.0	699	4 US-09-134-001C-4054	Sequence 4054, Ap
16	110.5	8.0	903	2 US-08-853-310-2	Sequence 2, Appli
17	110	7.9	413	1 US-08-700-749A-1	Sequence 1, Appli
18	110	7.9	413	3 US-09-020-684-1	Sequence 1, Appli
19	110	7.9	413	3 US-09-020-467-1	Sequence 1, Appli
20	110	7.9	413	3 US-09-020-685-1	Sequence 1, Appli
21	110	7.9	413	4 US-09-020-683-1	Sequence 1, Appli
22	109.5	7.9	1848	4 US-08-296-791-6	Sequence 6, Appli
23	109.5	7.9	1848	5 PCT-US95-10661A-6	Sequence 4, Appli
24	109	7.9	1104	4 US-08-923-992A-4	Sequence 6, Appli
25	109	7.9	1128	4 US-08-923-992A-6	Sequence 6, Appli
26	107.5	7.8	715	4 US-09-215-221-25	Sequence 25, Appli
27	107.5	7.8	1098	4 US-08-923-992A-8	Sequence 8, Appli

28	107	7.7	396	1 US-08-430-024-2	Sequence 2, Appli
29	107	7.7	396	1 US-08-782-009-2	Sequence 2, Appli
30	107	7.7	396	3 US-09-017-302-2	Sequence 2, Appli
31	107	7.7	1164	4 US-08-923-992A-2	Sequence 2, Appli
32	107	7.7	1702	4 US-08-296-791-5	Sequence 5, Appli
33	107	7.7	1702	5 PCT-US95-10661A-5	Sequence 5, Appli
34	105.5	7.6	1541	4 US-08-296-791-3	Sequence 3, Appli
35	105.5	7.6	1541	5 PCT-US95-10661A-3	Sequence 3, Appli
36	105.5	7.6	1545	4 US-08-296-791-4	Sequence 4, Appli
37	105.5	7.6	1545	5 PCT-US95-10661A-4	Sequence 4, Appli
38	104.5	7.5	219	4 US-09-134-001C-5651	Sequence 5651, Ap
39	104	7.5	788	2 US-08-918-914-4	Sequence 4, Appli
40	104	7.5	1099	4 US-09-442-100-2	Sequence 2, Appli
41	102.5	7.4	544	2 US-08-935-760-2	Sequence 2, Appli
42	102.5	7.4	664	1 US-08-421-661-6	Sequence 6, Appli
43	102	7.4	351	2 US-08-929-417-2	Sequence 2, Appli
44	102	7.4	930	4 US-09-134-001C-5314	Sequence 5314, Ap
45	101.5	7.3	542	1 US-07-814-964-13	Sequence 13, Appli

ALIGNMENTS

RESULT 1
US-09-599-652-2
Sequence 2, Application US/09599652
Patent No. RE37741
GENERAL INFORMATION:
APPLICANT: HOSTETTER, MARGARET K.
APPLICANT: GALE, CHERL A.
APPLICANT: BENDEL, CATHERINE M.
APPLICANT: TAO, NIAN-JUN
APPLICANT: KENDRICK, KATHLEEN
TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESS: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P. A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/599,652
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUEITING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1664 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-599-652-2
Query Match 100.0%; Score 1386; DB 1; Length 1664;
Best Local Similarity 100.0%; Pred. No. 2.8e-115;

	Matches	263;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	MNSTPSKLLP	IDKSHHLLQLOPOSSASSTFNSPTKPLNPRNTSKSLDNPSSDQTTYSQ	60						
Db	1	MNSTPSKLLP	IDKSHHLLQLOPOSSASSTFNSPTKPLNPRNTSKSLDNPSSDQTTYSQ	60						
QY	61	DOEKGEKKKDDTAFO	TSPEDRNFDLNSIDIQOTTIOHQOQOPOQOQOOLSGTDNNLLIDEEFSF	120						
Db	61	DOEKGEKKKDDTAFO	TSPEDRNFDLNSIDIQOTTIOHQOQOPOQOQOOLSGTDNNLLIDEEFSF	120						
QY	121	QTPMSTLDTLTKQNT	PTVYKVNENNAHPTIYNTPSKNSIMKKAKPKKSPKKAATVYNPEIH	180						
Db	121	QTPMSTLDTLTKQNT	PTVYKVNENNAHPTIYNTPSKNSIMKKAKPKKSPKKAATVYNPEIH	180						
QY	181	HYPDNRVVEEDDSQO	KEDSVEPRLIQHOQKDESGQENYSDSDTNAASVPRLPHTTKPTFA	240						
Db	181	HYPDNRVVEEDDSQO	KEDSVEPRLIQHOQKDESGQENYSDSDTNAASVPRLPHTTKPTFA	240						
QY	241	QLLNKNNEVNSEP	ALTDIMKLLK	263						
Db	241	QLLNKNNEVNSEP	ALTDIMKLLK	263						

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Db	1	MNSPTSKLLPDKSHHQLOQSSASATFNPSPTKPLNFPNRSKPSLDPMSSSOTTYISQ	60
Qy	61	DOEKREKKKOTAACTSPDRNFDLNDSIDIQOTIQHQOQOPQOQOOLQSDTDNNLIDEFST	120
Db	61	DOEKREKKKOTAACTSPDRNFDLNDSIDIQOTIQHQOQOPQOQOOLQSDTDNNLIDEFST	120
Qy	121	QTPMTSTDLTKONPTVDKVENHNAPTYINTSPKSIIMKKATPKASPKKAAFTVTNEIHH	180
Db	121	QTPMTSTDLTKONPTVDKVENHNAPTYINTSPKSIIMKKATPKASPKKAAFTVTNEIHH	180
Qy	181	HYPDNRVVEEEOOSOKEDSVPEPLIQHOMKPOPOFNYSDEPTNNAVPPPLHTTKRPTFA	240
Db	181	HYPDNRVVEEEOOSOKEDSVPEPLIQHOMKPOPOFNYSDEPTNNAVPPPLHTTKRPTFA	240
Qy	241	QLLNKNNVNSEPEALDMLDKLR	263
Db	241	QLLNKNNVNSEPEALDMLDKLR	263

```

RESULT 3
US-09-264-604-2
: Sequence 2, Application US/09264604
: Patent No. 6346411
: GENERAL INFORMATION:
: APPLICANT: HOSTETTER, MARGARET K.
: APPLICANT: GALE, CHERYL A.
: APPLICANT: BENDEL, CATHERINE M.
: APPLICANT: TAO, NIAN-JUN
: APPLICANT: KENDRICK, KATHLEEN
: TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
: TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MUETING, RAASCH, GERHARDT & SCHWAPPACH, P.A.
: STREET: 119 NORTH FOURTH STREET, SUITE 203
: CITY: MINNEAPOLIS
: STATE: MINNESOTA
: COUNTRY: USA
: ZIP: 55401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/264,604
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/542,846
: FILING DATE: 03-MAY-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: MUETING, ANN M.
: REGISTRATION NUMBER: 33,977
: REFERENCE/DOCKET NUMBER: 110.00280101
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-305-1217
: TELEFAX: 612-305-1228
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1664 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-264-604-2

Query Match 100.0%; Score 1386; DB 4; Length 1664;
Best Local Similarity 100.0%; Pred. No. 2,8e-115;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSTSKLLPDIKHSLOLOPOSSASJFNSPTKPLNFPRTNSKPSLDPNSSSDTYTSEQ 60

```


TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/264,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 236 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2
US-09-264-604-3

Query Match 17.4%; Score 241; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 4,1e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SDEDTNAYVPPPTPLHTTKPTFAQLLNKNNEVSEPALDMLKLR 263
|||||
Db 1 SDEDTNAYVPPPTPLHTTKPTFAQLLNKNNEVSEPALDMLKLR 46

RESULT 7
Sequence 3716; Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3716
LENGTH: 1335
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3716

Query Match 9.2%; Score 127.5; DB 4; Length 1335;
Best Local Similarity 21.2%; Pred. No. 0.006;
Matches 54; Conservative 41; Mismatches 125; Indels 35; Gaps 7;

QY 18 QLOPOSSASIFNSPTKPLNFPRTNSKPSLDPNSSDPTVT-SEDDQEKGEKKDTAFQT 76
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 32 QPOQNSHKRWLVDDQTLKQAEKAKSEVLTOSTNVSTQYQDPYQAPQDPQSTTYDA 91
QY 77 SPDRNFDLNSIDIOCTIOHQOOOPQOOQLSOTDNLLIDFES-OTPMSTLDTKONP 135
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 92 SLD-----EMSTYNEISSNKKOOSLSTDANONOTNSVTKNOOETNDLTOEDK 140
QY 136 TVQKVENHAPTYINTSPNSIMKKATPKKAVFTVNTPEIHHPDMRVEEED-QSQ 194
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 141 TSDITNLOETOSYAKENEDLGNANNNEQDKM--TASQPSENALETOTASNDNEQ 198
QY 195 QKEDSV-----EPPLIOHWKDPSPQFN SDEDTNAYVPPPTPLHTTKPTFAQLLNKN 247
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 199 QKSQVTSQENETATPKVSNMTNAGYNFDVDEDDSDSTHLEPISL-----N 246

QY 248 EVNSEPEALDMLKLR 262
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 247 NVNATSKOTTSYKXK 261

RESULT 8
US-09-134-001C-3159
Sequence 3159; Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3159
LENGTH: 10182
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 8.5%; Score 118; DB 4; Length 10182;
Best Local Similarity 21.2%; Pred. No. 0.67;
Matches 61; Conservative 45; Mismatches 112; Indels 70; Gaps 10;

QY 11 IDKRSHLQLOPOSSASIFNSPTKPLNFPRTNS----KPSLDPNSSDPTVTSEDDQEKGK 66
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 4526 INNESHILNNAQKDNFKAQVNSA-----PNHNLTETIKNRADTLNOSMTALSESIAIDYEN 4579
QY 67 EKKKDTAFQTSFDRNFDLNSIDIOCTIOHQOOOPQOOQLSOTDNLLIDFESQFQPMTS 126
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 4580 QKOQENTLDSNNKRRQDYDVAVNAKILNOTOSPTMSA-----DIVDQKADYKRTK 4632
QY 127 T-----IDLTKONP-----TVQKVENHAPTY--INTSPNSIKKATPKKASPKVA 171
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 4633 TALDGNORLEVAQOALNHLNTLNDLDAQROTLTDTINNSPINSVNAQKEKANTVNTA 4692
QY 172 FYTYNPEIHHYP-----NRVEEDSOQKEDSVPEPLIOHWKDP 212
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 4693 MTOLKQTIANYDELHDGYNINADKKDAYNNAVNNAKQLINQSDANOQAL-----DP 4746
QY 213 SDFNYSDEDTNAYVPPPTPLHTTKPTF--AQLLNKNNEVSEPALT 257
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 4747 AEIN-----KYTORVNTTKNDLNGDKLAEAKRDANTTIDGLT 4784

RESULT 9
US-08-923-992A-10
Sequence 10; Application US/08923992A
Patent No. 6280738

```

; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-1ga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923.992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-923-992A-10

Query Match      8.2%; Score 114; DB 4; Length 1164;
Best Local Similarity 18.5%; Pred. No. 0.08;
Matches 50; Conservative 52; Mismatches 80; Indels 88; Gaps 11;

Oy 22 OSSASIFNSPTKPLNFPRTNSKPSLDP-----NSSSDTYTSBDDQKKEKKD--- 71
Db 64 OGNNSSSELETKTIEIPTDIDKKAVERPLEKTAGETSAITDTGKRREKQLOQKNNLNKNDVH 123
Oy 72 -----TAFQTSFDRNFIDNSIDIOQT-----OHQO---QQPOQOQOLS 108
Db 124 NTLSHEQKNEFKTIDENDSDALLEENQENETNRLLHIKQHEVEKDKAKQOKTLK 183
Oy 109 QTD-----NLIDERSFQTPMTSTL-----DLTKONPTVD----- 138
Db 184 QSDTVVDLSNIDKELNHOQSQVETMAEQLGITNEDKDSMLKRIEDIRKQAOQADKKEAE 243
Oy 139 -KVNENHAPTYINTSP-----NKSIMKATPKASPKKAVATVYNTEIHHDN----- 185
Db 244 KAVRELGKLTSTKAGLDQETQIENHYKKTETSEENTQKVD-----EHTPSNLONLAQ 295
Oy 186 -RVEEDOSQEKDESVEPPLIQHOMKPSQ 214
Db 296 KSLFELDQATNEQATQ---VKNQFLENQ 322

RESULT 10
US-08-961-083-218
; Sequence 218, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-218

Query Match      8.1%; Score 112.5; DB 4; Length 565;
Best Local Similarity 22.0%; Pred. No. 0.041;
Matches 61; Conservative 37; Mismatches 102; Indels 77; Gaps 13;

Oy 1 MNSTPSKLLPIDKSHLQLOPOSSA--SIFNSPTKPLNFPRTNSKPSLDPNSSSTPTYS 58
Db 78 VESKPEKVAAPES---QPSDKPAESKVEQAGEPV-APREDEKAVPEP----- 123
Oy 59 EDDQEKREKKDIAFQTSFDRNFIDNSIDIOQTTOHQOQOPOO-----OOLSQTDNN 113
Db 124 EKQPEAREEKA-----VEETKQESSTDTTAEETVEKEETVQ 164
Oy 114 LIDFESFQTPMTSTDLTKONPTVDKVNENHAPTYINTSPNKSIMKATPKASPKKVAFT 173
Db 165 SIEQPVETPVAEKQTEPEEPKVEQAGEPVAPREDEQAP-----TAPVEPEK----- 212
Oy 174 VTNPELIHNPDRNVEEDOSQEKEDSV---EPLIQHOMKDPQSFNYSDEDTNASTPPT- 229
Db 213 --QPEVPE--EKAVAETTPPEDEKIGIKTPE-----VDKSELN-NOIDKASVSPTD 261
Oy 230 -----PLHTTKPTPEAQLLNKNNEVNSPEAL 256
Db 262 YSTASYNALGPVLETKAGYVASEPVKQPEVNSETNKL 298

RESULT 11
US-08-242-932-2
; Sequence 2, Application US/08242932
; Patent No. 5595740
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 5595740-1ga Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
```


STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06111
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/242,932
FILING DATE: 16-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Salliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI42
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-06111-2

Query Match	8.18;	Score 112;	DB 5;	Length 984;
Query Match	8.18;	Score 112;	DB 5;	Length 984;

Matches 56; Conservative 54; Mismatches 115; Indels 88; Gaps 11;

```

0y 22 QSSASIAINSPTKPLNFPRTSKSLDP-----NSSDPYTSSEOQOENKKEKRO--- 71
Db 27 QGNSSSSSELETTKMEIPTTIDIKKAVEVEVEKTAGETSATPDGKREKLOQOKNNKLKNDVD 86
0y 72 -----TAFQTSFDRNFDLNSIDIQOTI-----QHOQ---QOPQOQOOLS 108
Db 87 NTIISHEOKNEFKRIEOTNDSDLLELENOFNETNRLLIHIOHEEKEKDKKAQOOKTLK 146
0y 109 QTD-----NNEIDFESFQTPMSTSL-----DLTKONPTVDKVENH 144
Db 147 QSDFRVDSLNSIDKELNHQKSOVEKMAEQKGTINEDKOSMLKIEDIKQAQOQDKKDAE 206
0y 145 APTIYNTSPKNSIMKKATPKKASPPKVAFTYVNPETIHHYPDNKRVDEE-----DOSQOK-EDS 199
Db 207 VKVOLEEBEASHKLQAVEDFRKKKRTSEQV-----PKKRVKRDILAANNENNOOKTELT 259
0y 200 VEPPLIOHQMKDPSOFNSYDDEFTNAAVPPPEPLHTT-----KPFQAQLLNKNEV 249
Db 260 VSPENIT-----VEGEDVAFVTYAKAKSDKTTIIDFSDLLTKYMPVSDRSTRYKTI 310
0y 250 NSEPEALTDMLK 262
Db 311 NTFDNHKAETIETIK 323

```

RESULT 14
US-09-134

```

: Sequence 3511 Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964

```

```

? PRIOR FILING DATE: 1997-11-08
? PRIOR APPLICATION NUMBER: US 60/055,779
? PRIOR FILING DATE: 1997-08-14
? NUMBER OF SEQ ID NOS: 5674
? SEQ ID NO 3517
? LENGTH: 652
? TYPE: PRT
? ORGANISM: Staphylococcus epidermidis
? US-09-134-001C-3517

```

Query Match	8.08; Score 111; DB 4; Length 652;
-------------	------------------------------------

Matches 49; Conservative 26; Mismatches 57; Indels 52; Gaps 10;

Qy	12	DKSHILOPOSSASISINSPKPLINPRITMSKS-----LDNNSSDPITSODD-EK	64
Db	94	ENOSGNVOKSNQIODDSTOTSPINDOKQTSMEQSKDNHVTNRSODTYPKGONODDK	155
Qy	65	GKEEKKD-----TAFQ-TSPDRNFDLNSIDLOOTIGHOO-----OPQOOOOLSDTD	111
Db	154	GKQFKQKDNHQSOTGQNPQTQONNDQDSS-DKKHNPSPQTDSSSKGPRQOSIEDRD	212
Qy	112	NNLIDESFGTPMTSTDLTKQNP--VDKYNENHAPTYINTSPKSI--MKKATPKAS	166
Db	213	K-----TVKOPSSKVNKIG-----NTKTDKTVTKQKOTSLTS	246
Qy	167	PKKV 170	
Db	247	PRVV 250	

RESULT 15

```

US-09-134-001C-4054
: Sequence 4054, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 4054
: LENGTH: 699
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4054

```

Query Match	8.08;	Score 111;	DB 4;	Length 699;
-------------	-------	------------	-------	-------------

Matches 53; Conservative 37; Mismatches 103; Indels 38; Gaps 10;

```

0Y  11 IKDHSHLLODPOSSAS1-----FMSPPKPLNFPNTNKPESLDNPNSSSDYTSQDOBK 65
Db  134 VAKHS--BEKPOOEVELEKHAHSENNOTLHSAKAOSNDYVTKTSQOLDNTTAQOEDSOKE 199
0Y  66 KEERKDTAFOTSFD--RNFDLDNSIDIOQTIO-HOQOQPOOQOOLQSDNNLIDIEFSOT 122
Db  192 NLSKQDQTSKRTDLLRATGQNSKDSOSTEEVKNKVAKDQYAKAKDDQVETPLNS 25
0Y  123 PMTSTLDTLTKO-NPTVDKVYENHNAFTYINTSPNKSIMKATPKASPKKVAFTVNPETIH 189
Db  252 -KEEPLKDKQKQNPPTDK-----DKSKNDKGSHDGLANLESNAVAAT----- 293
0Y  182 YDNNVVEEDDSQOKEDSEVERPLLOHQMK--DP-----SQFYSDBDNTASV 226
Db  294 --NQSQOQVESEKNEQDTNKSAKQOYKRNNDPILVLHGFGFTDINPVS 341

```

Mon Mar 17 12:50:24 2003

us-09-964-858-1_copy_1_263.ra1

Page 8

Search completed: March 17, 2003, 12:27:47
Job time : 17.0944 secs

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:13:06 ; Search time 3.27556 Seconds
(without alignments)
3330.204 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263

Perfect score: 1386
Sequence: 1 MNSTPSKLLPIDKHSHLQ.....NKNNEVNSEPEALTDKLR 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1386	100.0	1664	1 INT1_CANAL	P53705 candida alb
2	148	10.7	971	1 CL4A_CANAL	O14427 candida alb
3	141.5	10.2	236	1 LIPB_MYCPU	Q91806 mycoplasma
4	132	9.5	499	1 WETA_PENCH	O01870 penicillium
5	130.5	9.4	1380	1 ZMS1_YEAST	P46974 saccharomyc
6	130	9.4	1087	1 POM1_SCHPO	Q09699 schizosacch
7	128.5	9.3	1159	1 N124_SCHPO	Q09904 schizosacch
8	125	9.0	309	1 YJ53_YEAST	P47129 saccharomyc
9	123	8.9	1139	1 HMW1_MYCGE	Q49413 mycoplasma
10	123	8.9	1616	1 P200_MYCGE	Q49429 mycoplasma
11	121	8.7	605	1 YHC8_YEAST	P38739 saccharomyc
12	120.5	8.7	910	1 DNMU_MYCPN	P75354 mycoplasma
13	120.5	8.7	916	1 TOP1_ARATH	P30181 arabidopsis
14	119.5	8.6	599	1 HMW3_MYCGE	Q57081 mycoplasma
15	119.5	8.6	872	1 SCDS_YEAST	P34758 saccharomyc
16	119	8.6	574	1 MIG1_CANAL	O97792 candida alb
17	118	8.5	842	1 CHS5_SCHPO	P48563 saccharomyc
18	116	8.4	620	1 CHS5_YEAST	Q92357 schizosacch
19	116	8.4	1023	1 CLOC_DROME	O61733 drosophila
20	116	8.4	1231	1 YKT3_CAEEL	P10354 caenorhabdi
21	115.5	8.3	466	1 CMGA_RAT	P10354 rattus norv
22	115	8.3	198	1 HUNB_DROLI	O46244 drosophila
23	115	8.3	700	1 B1B_DROME	P23645 drosophila
24	115	8.3	925	1 P1P1_YEAST	P40020 saccharomyc
25	114.5	8.3	648	1 KAPC_DICDI	P34099 dictyosteli
26	114.5	8.3	701	1 CGL_HUMAN	Q13495 homo sapien
27	114.5	8.3	1083	1 YTL2_YEAST	P40480 saccharomyc
28	114.5	8.3	1177	1 SP97_DICDI	O95293 dictyosteli
29	114.5	8.3	1365	1 SUZ2_DROME	P25172 drosophila
30	114	8.2	779	1 MSS4_YEAST	P38994 saccharomyc
31	114	8.2	850	1 D7_DICDI	P34682 dictyosteli
32	114	8.2	1113	1 MG21_YEAST	P40578 saccharomyc
33	113.5	8.2	448	1 TANK_MOUSE	P70347 mus musculu

34	113.5	8.2	2469	1 TEGU_HSVSA	O01056 herpesvirus
35	112.5	8.1	583	1 APL_KLUFA	P56095 kluyveromyc
36	112	8.1	612	1 ADPL_CANAL	P46589 candida alb
37	111.5	8.0	896	1 CBL_MOUSE	P22682 mus musculu
38	111	8.0	1381	1 YHE7_YEAST	P34216 saccharomyc
39	111	8.0	1694	1 IGA0_HAEIN	P44969 haemophilus
40	110.5	8.0	705	1 DP02_YEAST	P38121 saccharomyc
41	110.5	8.0	847	1 DNLI_CAEEL	O27474 caenorhabdi
42	110	7.9	656	1 P13_MOUSE	O60865 mus musculu
43	110	7.9	688	1 LIP_STAEF	O02510 staphylococ
44	110	7.9	1110	1 TRAI_CAEEL	P34708 caenorhabdi
45	110	7.9	1341	1 YL78_YEAST	O05854 saccharomyc

ALIGNMENTS

RESULT 1	ID	INT1_CANAL	STANDARD	PRT: 1664 AA.
AC	P53705	INT1_CANAL		
DT	01-OCT-1996	(Rel. 34, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Integrin alpha chain-like protein (Alpha-INT1).			
GN	INT1			
OS	Candida albicans (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
OX	NCBI_TaxID=5476;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 10261;			
RX	MEDLINE=96133936; PubMed=8552638;			
RA	Gale C., Finkel D., Tao N., Meinke M., McClellan M., Olson J.,			
RT	Kendrick K., Hostetter M.;			
RT	"Cloning and expression of a gene encoding an integrin-like protein			
RT	in Candida albicans.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:357-361(1996).			
CC	-!- FUNCTION: COULD PLAY A ROLE IN ADHESION AND IN STE12-INDEPENDENT			
CC	MORPHOGENESIS.			
CC	-!- SUBCELLULAR LOCATION: CELL-SURFACE OF THE BLASTOSPORES.			
CC	-!- SIMILARITY: SOME, WITH INTEGRIN ALPHA CHAINS.			
CC	-!- SIMILARITY: CONAINS 1 PH DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U35070; AAA96019.1; -			
DR	InterPro: IPR001849; PH.			
DR	Pfam: PF00169; PH: 1.			
DR	SMART: SM00233; PH: 1.			
DR	PROSITE: PS50003; PH_DOMAIN: 1.			
KW	Cell adhesion; Glycoprotein.			
FT	DOMAIN	1527	1636	PH.
FT	SITE	1149	1151	
FT	CARBOHYD	50	50	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	154	154	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	216	216	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	223	223	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	265	265	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	268	268	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	285	285	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	289	289	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	306	306	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	314	314	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	446	446	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	583	583	N-LINKED (GLCNAc. . .) (POTENTIAL).

```
FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 838 838 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 841 841 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1018 1018 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1082 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1100 1100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1113 1113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1200 1200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1571 1571 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1593 1593 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 95 106 POLY-GLN.
FT DOMAIN 283 286 POLY-ASN.
FT DOMAIN 1283 1292 POLY-HIS.
FT DOMAIN 1651 1660 POLY-GLN.
SQ SEQUENCE 1664 AA: 187859 MM: BAEFDDDC8196790 CRC64;

Query Match 100.0%; Score 1386; DB 1; Length 1664;
Best Local Similarity 100.0%; Pred. No. 4, 2e-77;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSTPSKLPIDKSHLOPOSSASISFSPKPLNFPRTNSKPSLDPMSSSDPTTSEQ 60
DB 1 MNSTPSKLPIDKSHLOPOSSASISFSPKPLNFPRTNSKPSLDPMSSSDPTTSEQ 60
QY 61 DQKGRKEDTAFOTSFDRNFIDNSIDIQOTIOHQOQOPOOQOQOQSDNDNMLIDFESF 120
DB 61 DQKGRKEDTAFOTSFDRNFIDNSIDIQOTIOHQOQOPOOQOQOQSDNDNMLIDFESF 120
QY 121 QTPMTSTLDTKONPTVDKVENHAPTYINTSPKNSIMKRAKPKKVAFTVTNPEIH 180
DB 121 QTPMTSTLDTKONPTVDKVENHAPTYINTSPKNSIMKRAKPKKVAFTVTNPEIH 180
QY 181 HYBDNRVEEDSOQKEDSVPEPLIOHQMDPSOFNYSDEDTNNAVPTPTPTKTFPA 240
DB 181 HYBDNRVEEDSOQKEDSVPEPLIOHQMDPSOFNYSDEDTNNAVPTPTPTKTFPA 240
QY 241 QLNKNNEVSEPEALTDMKLK 263
DB 241 QLNKNNEVSEPEALTDMKLK 263

RESULT 2
CLA4.CANAL STANDARD: PRT: 971 AA.
15-JUL-1998 (Rel. 36, Created)
15-JUL-1998 (Rel. 36, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Serine/threonine-protein kinase CLA4 (EC 2.7.1.1.-).
GN CLA4.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_Taxid=5476;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97411146; Pubmed=9259554;
RA Leberer E., Ziegelbauer K., Schmidt A., Marcus D., Dignard D., Ash J.,
RA Johnson L., Thomas D.Y.;
RT "Virulence and hyphal formation of Candida albicans require the
RT Ste20p-like protein kinase Cla4p."
RL Curr. Biol. 7:539-546(1997).
CC -1- FUNCTION: ESSENTIAL FOR VIRULENCE AND MORPHOLOGICAL SWITCHING
CC (HYPHAL FORMATION) OF C. ALBICANS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U8796; AAB6813.1; -.
DR HSSP: Q63450; 1A06.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR000095; Pakbox/Rhodning.
DR InterPro: IPR001849; PH.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00786; PBD; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00285; PBD; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS50108; CRIB; 1.
DR PROSITE: PS50003; PH DOMAIN; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Transferrase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 67 178 CRIB.
FT DOMAIN 231 244 PH.
FT DOMAIN 680 935 PROTEIN KINASE.
FT DOMAIN 16 19 POLY-SER.
FT DOMAIN 210 213 POLY-PRO.
FT DOMAIN 311 315 POLY-ASN.
FT DOMAIN 384 389 POLY-ASN.
FT DOMAIN 392 395 POLY-ASN.
FT DOMAIN 418 421 POLY-PRO.
FT DOMAIN 453 462 POLY-GLN.
FT DOMAIN 468 476 POLY-GLN.
FT DOMAIN 572 576 POLY-GLN.
FT NF_BIND 686 694 ATP (BY SIMILARITY).
FT BINDING 710 710 ATP (BY SIMILARITY).
FT ACT_SITE 803 803 BY SIMILARITY.
SQ SEQUENCE 971 AA: 106889 MM: ADF5DDBCC6CF624B CRC64;

Query Match 10.7%; Score 148; DB 1; Length 971;
Best Local Similarity 24.1%; Pred. No. 0.041;
Matches 77; Conservative 39; Mismatches 100; Indels 104; Gaps 15;

QY 4 TPKSLPDKSHLOPOSSASISFSPKPLNFPRTNSKPSLDPMSSSDPTTSEQ 40
DB 358 TPYHLTQLNGSSH---QHTSSGSLPSSGNNNNNSTNNNTKAVSPILNLMKSELLPA 414
QY 41 TNSKPSLDPMSSSDPTTSEQDQKGRKEDTAFOTSFDRNFIDNSIDIQOTIOHQOQO 100
DB 415 RRAPEPTSGTSDPTTSEKKNHIDNSGYEQ-----KQQRDSSQOQOQOQHORYOQ 465
QY 101 PQOQOQOQSDNDNMLIDFESFQTPMTS---TLDTKONPTVDKVENHAPTYINTSPKNS 156
DB 466 KSOQOQOQO-----QQLPSHQGTSHIPKQVP-----PTLPSSGPP-- 502
QY 157 IMKATPKKASKKVAFTVTNPEI-----HHY-----PDNRVEE 190
DB 503 -TQAASGKSMPSKI-----HPDLIQGTNNYIKSSGTDANQVGDAAKQKIPFNLSQSK 556
QY 191 DOSQKEDSVPEPLIOHQMDPSOFNYSDEDTNNAV-----PPNPPLHTTTKPTFAOL 242
DB 557 SQOQLASKQSPSSQOQOQKP-WTSHGLMGTSHTSVKPLNPVNDPIKPLNLKSKSKEA 615
QY 243 LKNKNEVSEPEALTDMKLK 262
DB 616 LNFTSGVSKTSP-TDKSNK 634
```

```

RESULT 3
LIPB_MYCPU          STANDARD;      PRT;       236 AA.
AC Q9LRP6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein B precursor.
GN LIPB OR MYPU_5220.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT.
RA MEDLINE=20245550; PubMed=10781561;
RT "Gene rearrangements in the ysa locus of Mycoplasma pulmonis.";
RL J. Bacteriol. 182:2900-2908(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RA MEDLINE=21267165; PubMed=11353084;
RT Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Gailsson F., Moszer I., Dybvig K., Wroblewski H., Viari A., Kocha E.P.C., Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis.";
CC Nucleic Acids Res. 29:2145-2153(2001).
CC -! SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (potential).
CC -! SIMILARITY: BELONGS TO THE PULMONIS LIPAB LIPOPROTEIN FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/cc or send an email to license@isb-sib.ch).
-----
DR EMBL; AF198037; AAF70133.1; -
DR EMBL; AL445564; CAC13695.1; -.
DR Mypulist; MYP_U_5220; -.
KW Lipoprotein; Membrane; Signal; Complete proteome.
FT SIGNAL                     1         27    POTENTIAL.
FT CHAIN                      28        236   LIPOPROTEIN B.
FT LIPID                       28        28    N-ACTYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN                     82       158    ASN-RICH.
TT SEQUENCE                  236 AA; 26661 MW; 183DIB968CDC1D3A CRC64;
-----
Query March                               10.2%; Score 141.5; DB 1; Length 236;
Best Local Similarity 21.3%; Pred. No. 0.021; Matches 49; Indels 49; Gaps 7;
Matches 47; Conservative 49; Mismatches 76;

QY 52 SSDVTYSDEODEKKEE-----KKDTATGTSDFRNFEDLDNSIDIOQTIOHQOQDPQQ 103
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 26 ASCSYNLAKERKDQDKESTNLSEPKMSUTSKTNFQDKDKDSYNTKIDSQESSKTSQSNTSE 85
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 104 QQQLSDTDNNLLDEFSEFOTPMTSTILDITLKONFTYDKVN---ENHAPTYINTSPKSIKK 160
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 86 SNQNKTKVDS-----KTNNLATQNQNPSSKNVINOETNETRKQEOINIPNNAVISSE 135
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 161 ATPKSPKVAVFPTVYNPEIINHYPDNKRVEEEDSOOKEDSVEPPLQLQHMKDPSQNYSDSE 220
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 136 --KQENDKNASSLSKOI-----NNTLTNODTTCOE-----OFKQESK 174
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 221 DTNASVPRTPLH-----TTKPPFAOLLNKNNEVNSEPE 254
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 175 DSNNTSPKPIITHDINKIVISSQSTRLEMPKNDQSNSESE 215

```

[illegible]

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Zinc finger protein ZMS1.
GN ZMS1 OR YJRI27C OR J2052.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=RC11-6A.
RA Thomas D., Bardey R., Surdin-Kerjan Y.;
RL Submitted (xxx-1993) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA Rose M., Koeltter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- SIMILARITY: STRONG, TO YEAST YML081W.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L26506: AAA3240.1: -
CC EMBL: Z49627: CA89658.1: -
CC HSSP: P07248: 1ARD.
CC TRANSFAC: T04595: -
CC SGD: S0003888: ZMS1.
CC InterPro: IPR000822: znf_C2H2.
CC Pfam: PF00096: zf-C2H2; 2.
CC PRINTS: PR00048: ZINCINGER.
CC SMART: SM00355: znf_C2H2; 2.
CC PROSITE: PS00028: ZINC_FINGER_C2H2_1; 2.
CC PROSITE: PS00157: ZINC_FINGER_C2H2_2; 2.
CC DNA-binding: Nuclear protein; zinc-finger; Metal-binding.
FT DOMAIN 64 67
FT ZN_FING 151 173
FT ZN_FING 179 202
FT DOMAIN 424 433
FT CONFLICT 1 17
FT CONFLICT 1117 1118
FT CONFLICT 1131 1131
FT CONFLICT 1143 1380

SQ SEQUENCE 1380 AA: 155062 MW: 155062 MD: FBE987EAB5B639000 CRC64;
(IN REF. 1).

Query Match 9.4%; Score 130.5; DB 1; Length 1380;
Best local Similarity 18.9%; Pred. No. 0.7;
Matches 66; Conservative 53; Mismatches 121; Indels 109; Gaps 13;

QY 3 STPSKLLPDKHSHLQLOPOSSASIFNSPTKPL-NPRTNKSRLDPNSSSDTYTSEOD 61
DB 331 NTPSSMHTKRHASFSSAMTYMSSNSPHHSITNELVEDAPH-QVGFSTPMTAKOL 389
QY 62 QEKGEK-----EKDTPAFQTSFDKN-FDLNLSIDIOQTQHQQQPOQQQQLSOTDN 112
DB 390 MESVSELDLPPLTLDEPQAIK--FNLTLPNDPS-----GQQQQQQQQQQNSTSS 438
QY 113 NLIDEFQTPMTSTLDLTQNPFTVK-----VNEHNP----- 146
DB 439 TIYNSNNGSVATIPGVYLLSSGFSLDTLTMNSAHAGAGGYMSSHSHSPDLGCRSHDKPT 498

QY 147 -----TYINTSP-----NKSIMKKATPKASPKVAFTVTN 176
DB 499 VSEFNLSPSPNTIPSNSTTASNSYLNANQTYROMNEQPLMSLSPKNPPTVSDSSST 558
QY 177 PEIHHPDNKREEDSOQAKEDSEVPELLOH-----KDPQSO-----FN 216
DB 559 INFNGTNNLLEPSEMPNDNSIDPAIDKMLSEFINNSDPSTKINFENHNDIGFI 618
QY 217 YSDEDTNAVSP-PPPLHTT-----KPPPAOLLNKNEVNSEPE 254
DB 619 YSPSSRSSIIPKSPNHSATSLNHEKASLSPRLNLSNGSTDLPSITPQ 667

RESULT 6
POM1_SCHPO
ID POM1_SCHPO STANDARD; PRT; 1087 AA.
AC 009690:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative dual specificity protein kinase pom1 (EC 2.7.1.-).
GN POM1 OR SPAC2F7.03C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean C.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voickert G., Aert R., Roden B., Gymnopoulos B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moser D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eber P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shipkowsky G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880(2002).
RN (2)
RP CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=98241493; PubMed=9573052;
RT Baehler J., Pringle J.R.;
RT "Pom1, a fission yeast protein kinase that provides positional
RT information for both polarized growth and cytokinesis";
RL Genes Dev. 12:1356-1370(1998).
CC -1- FUNCTION: INVOLVED IN LOCALIZATION OF POLARIZED GROWTH AND
CC CYTOKINESIS. MAY INTERACT WITH BOTH THE ACTIN AND MICROTUBULE
CC CYTOSKELETON. REQUIRES TEAL FOR LOCALIZATION TO THE CELL ENDS BUT
CC NOT TO THE CELL CENTER.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MNB/DYRK SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```

OY 179 IHVYDNRVREEDOSO-----OKEDSVPEPLIOHWKDRPQFVNSD----- 219
Db 430 FHHQPLSKOTEPKAKQFSSPTKESTTKSEVEPPSPKEIKS-SHFSVPEKFEPPKEA 488
OY 220 -EDTNAVSP-----PTPLP-----HTTKPT-FAQLLNKNVNS-----EPEALTD 259
Db 489 TWDKLNVKPEFKPTATVADVQTNRLKEMEPKPTFFAQLPSKTOGTPSTTEKKPFFSGL 548
OY 260 KLRK 263
Db 549 SPKR 552

RESULT 8
V53_YEAST
ID V53_YEAST STANDARD: PRT: 309 AA.
AC P47129;
RT 01-FEB-1996 (Rel. 33, Created)
01-FEB-1996 (Rel. 33, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
Hypothetical 35.5 kDa protein in MIR1-SHE18 intergenic region.
YXR083C OR J1857.
CN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA MEDLINE=96437976; PubMed=8840504;
RT Huang M.-E., Manus V., Chuat J.-C., Galibert F.;
RT "Analysis of a 62 kb DNA sequence of chromosome X reveals 36 open
RT reading frames and a gene cluster with a counterpart on chromosome
RT XI.";
RL Yeast 12:869-875(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 249583; CAA89610.1; .
CC DR EMBL: L47993; AAB39306.1; .
CC SGP: S0003843; YXR083C.
CC Hypothetical protein.
CC SEQUENCE 309 AA; 35520 MW; FE29EBD0192B2113 CRC64;

Query Match 9.0%; Score 125; DB 1; Length 309;
Best Local Similarity 22.5%; Pred. No. 0.29;
Matches 64; Conservative 36; Mismatches 124; Indels 60; Gaps 10;

OY 7 KLLPDKHSHLD-----LDQSSASIFNSPTKPLNPRTRNSKPSLDPNSSSDTY 56
Db 16 KLSIVDKHSGOQQOQPHQKQHEVQESKSPRV-----TPPLK-PKRLAIRISSPQSTNQ 70
OY 57 TSEDOQEKKEKKRTAFTGTSFDRNFDLNSIDIOQTIOHQOQOPOQOQOLSQTDNNLID 116
Db 71 SPVSHASPISTDOOLIKYLAAKHREINELSKLEVA-----OKELQOLEL---- 116
OY 117 EFSFGPMTSLDLTKONPT--VDKVNENHATY--INTSPKSIKTKATPKASPKVAF 172
Db 117 QKQDPLPRKGOQKLGQNFSEYLSFTKTIQOTFVDVNNSPMLKGGKSINDFEK----- 172
OY 173 TWVNEPEIHVYDNRVREEDOSOQKEDSVPEPLIOHWKDRPQFVNSDDEDTNAVSP-TPP 231
Db 173 -----PNNVNSNNTNLTLPNKKRPVPRNSQRMQNLAPSRSSSESTPTSGPRLLP 222
OY 232 LHTTK-----PTFAQLLNKNVNSPEPALDMLKLR 263

```

```

DB 223 RNTMKNANTTATAGENTPELQRIILKNFMQNMMEDEDFDLLEKR 266

RESULT 9
HMM1_MYCGE STANDARD: PRT: 1139 AA.
ID HMM1_MYCGE Q49413; Q49365;
AC Q49413; Q49365;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytoplasmic high molecular weight protein 1 (Cytoplasmic accessory
DE protein 1).
DE HMM1 OR MG312.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gockyne J.D., White O., Adams M.D., Clayton R.A.,
RA Fritschman J.R., Bult C.J., Kariya A.R., Sutton G., Kelley J.M.,
RA Fritschman J.R., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Ultebeck T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Luchter T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RN science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 721-847 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL j. Bacteriol. 175:7918-7930(1993).
CC -I- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC -I- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC MEMBRANE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isg.slb.ch/announce/
CC or send an email to license@slb.slb.ch).
CC -----
CC EMBL: U39712; AAC71534.1; -.
CC DR EMBL: U02261; AAD12527.1; -.
CC DR TIGR: MG312; -.
CC KW Cytoplasmic; Structural protein; Complete proteome.
CC SEQUENCE 1139 AA; 130531 MW; 0011D3286C3DB856 CRC64;

Query Match 8.9%; Score 123; DB 1; Length 1139;
Best Local Similarity 24.6%; Pred. No. 1.6;
Matches 66; Conservative 39; Mismatches 103; Indels 60; Gaps 15;

OY 11 IDKHS---HLDQRPSSSASIFNSPTKPLNFRPTSKPSLDPNSSSD--TYSDDQDEKG 65
DB 228 VDPSSDHFAPKIPST-----TDSYSPSDLRQLPRLDPSLDHNVQYDFNDHEELK 279
OY 66 KEKKDQTAFOFSFDR-NFDLNSIDIOQTIOHQOQOPOQOQOQLOSDTNMLIDEFSGOTPM 124
DB 280 PVAEEQNNYQVGFQDVQANLDNNEIDLPRAEKAYVTTPRESKQAOVVD-----SYGLRI 332
OY 125 -TSLIDL-----KONPTV---DKVENENHAPTYINTSPNKSIMKATPKASPK-----VA 171

```

Db 333 DTDDQDTFFSSFEPTQVDFQVNSB-----VNDQFKPEITKEPVLESSFNKQDVE 387
QY 172 FTVPNPEIHHYPDNRVEEEDSOQKEDSEVPEPLIOHMKDPDSQFNDDTASV---PP 228
Db 388 TSDLNSESNLYSEN-----NKDATNNDSDNSEFT-----QLNSNSETASDVHESK 434
QY 229 TPELHTTKPTFAQLL---NKNNEVNSEP 253
Db 435 SEPIDHYK--FGSDLSQSNNSNLSLESEP 460

RESULT 10
P200_MYCGE STANDARD: PRT: 1616 AA.
ID Q49429; Q49298; Q49352; Q49353;
AC 02-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein P200.
GN MG386.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
Fleischmann J.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
Fritchman J.L., Weidman J.F., Sandusky M., Fuhman J.L.,
Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
Tombl J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
"The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -I- FUNCTION: COULD BE A ACCESSORY STRUCTURAL COMPONENT IN
CYTADHERENCE (BY SIMILARITY).
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC
CC EMBL: U02245; AAC71613.1; -
DR EMBL: U02245; AAC71613.1; -
DR EMBL: U02245; AAC71613.1; -
DR EMBL: U02175; AAD12458.1; -
DR EMBL: U02126; AAD12402.1; -
DR EMBL: U02126; AAD12402.1; -
KM TIGR: MG386; -
DR Cytaadherence; Structural protein; Repeat; Complete proteome.
FT DOMAIN 1205 1389 2 X 32 AA REPEAT.
FT REPEAT 1205 1389 1-1.
FT REPEAT 1358 1389 1-2.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 891 1389 2-1.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT REPEAT 256 256 P -> S (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
SQ SEQUENCE 1616 AA; 185678 MW; 6AF76A13AC49E4FF CRC64;

Query Match 8.9%; Score 123; DB 1; Length 1616;
Best Local Similarity 20.4%; Pred. No. 2.4;

Matches 49; Conservative 52; Mismatches 111; Indels 28; Gaps 9;
QY 4 TPKSLPLIDKHSHL---QIQP-QSSASITFNSPTKRLNPRNRSKSL-----DP 49
Db 1202 TQPEIQVESPQEPAPFDVQVQPTQPEAKFSDSEVEEQEPSPSSSEPTQOHVESEASFDSP 1261
QY 50 NSSDPYTSSEQDQEKKEKKDTAFQTSFDR-NFIDLN-SIDIQITIQHQOQOQOQOOL 107
Db 1262 NTFDEPNNDPFDQPSYSDSLQSPSPQYDVERHYDDEPNYELSKSPSPQEPQVEQOP 1321
QY 108 SOTDNNLIDFESFQTPMTSTLDITKONPTVDKVENHAPTYINTSPKSIKKAIPKASP 167
Db 1322 GEAVERPSAEAKFSDSEVESVQD-SQEPILLEEVQTPQPEIQVESPQ-EATFTDPEQTPR 1379
QY 168 KQVAF-----TYTNPEIHHYPDNRVEEEDSOQKEDSEVPEPLIOHMKDPDSQFNDDTASV 222
Db 1380 QEAKFDSFVETIQEPQVSSSEPEVVOQPNFEKKPFTVLEP-----QADEIQPEASEES 1434

RESULT 11
YH08_YEAST STANDARD: PRT: 605 AA.
ID YH08_YEAST
AC P38739;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 63.8 kDa protein in GUT1-RIM1 intergenic region
DE precursor.
GN YH028W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Gattung S., Gelsel C., Kirsten J.,
Du Z., Favell A., Fulton L., Gattung S., Gelsel C., Kirsten J.,
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y., Mouser L.,
Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Vaughan K.,
Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
Vaughan M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RL Science 265:2077-2082(1994).
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC
CC EMBL: U11583; AAB65040.1; -
DR PIR: S48940; S48940.
DR SGD: S0001020; WSC4.
DR InterPro: IPR002889; WSC.
DR Pfam: PF01822; WSC.1.
DR SMART: SM00321; WSC.1.
KM Hypothetical protein; Transmembrane; Signal.
FT SIGNAL 1 26
FT CHAIN 27 605
FT DOMAIN 116 317
FT TRANSMEM 415 435
FT TRANSMEM 415 435
SQ SEQUENCE 605 AA; 63807 MW; C9DF8C8AA9553811 CRC64;

Query Match 8.7%; Score 121; DB 1; Length 605;
Best Local Similarity 22.9%; Pred. No. 1.1;
Matches 65; Conservative 32; Mismatches 85; Indels 102; Gaps 10;

```

OY 21 POSSASIFNSPTKPLNTPRTKSKSLDPNNSSPOTYSDEOQEGKEKKOTAFOQTSFDR 80
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 266 P7SSAPPTSSSNSTPTSTPTSTPTTSTSTAPSSSTLYATYS -----TTASP 310C
OY 81 NFDLDNSIDIQOTIOHQOQOQPOQOQOOLSCOTNNLIDFESFOTPMTSLDCLKOMPVDKV 140
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 311 ITSTTTSVNLQTSLK -----SVITVSVHTMOTNTSEITSRLTLPKKV 355
OY 141 -----NENHAPTYINTSPKXSIMKKTATPRASKV 170
Db 355 ITQIYSSLTGATPTSAVATTSASVGRITNNNSNTTNSNTPTPKSTEEKGY-WDSGCKI 413
OY 171 AFT-----VTNPEIHHPDNRVVEEDQSQOKEDSVEPPLIQHQMKDPSQFN 217
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 414 AATFVVGVCVLCIILCIILYILHHY-----RTRPARKAQOFENEGYSKQYQSK-----Y 462
OY 218 SDEDTNASVPTPLHTTKKPTFAOLKNKNNEVNSEPAL-TDMK 260
Db 463 PNEYV-----TTTLHTHTSPS-----SNSTSTPRLIYDEK 493

```

```

RESULT 12
DNJM_MYCPN          STANDARD:      PRT:      910 AA.
ID   DNJM_MYCPN
AC   P75334;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DnaJ-like protein MG200 homolog.
GN   MPN19 OR MP035.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RA   MEDLINE=97105885; PubMed=8948633;
RA   Himmelfreidh R., Hildert H., Pliagens H., Plickl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -1 SIMILARITY: CONTAINS 1 J DOMAIN.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL; AE000004; AAB95683.1; -.
DR   HSSP; P25685; 1HDJ.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ_1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS00636; DnaJ_1; 1.
DR   PROSITE; PS50076; DnaJ_2; 1.
KW   Hypothetical protein; Chapterone; Complete proteome.
LT   DOMAIN
TT   SEQUENCE 910 AA; 100190 MW; 125DDE37D2D221A7 CRC64;
SO   Query Match 8.7%; Score 120.5; DB 1; Length 910;
Best Local Similarity 21.6%; Pred. No. 1.8;
Matches 62; Conservative 52; Mismatches 106; Indels 67; Gaps 15;

QY 2 NSTPSKLLPIDKSHSLQLOPOSSASIFNSPTKPLNFPRTNSKPSLDPNSSDYVTSQD 61
   |||  -:-:  ::|||:-  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:
DB 294 NSAPF---PVDVETPEVLEQPEPEPITLSEPEAP---ASVVIETPLPELEERTSAVE 347
   :  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:
QY 62 QEKQK-KEEKKDTAFQTSDFRNFEDLDNSIDIQQTQHQQQQLPQQQQQLSQTQNNLIDERSF 120
   :  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:  -:-:

```

```

Db      348 MDASVKAADVDEA-----DATNEPTBEDTL-----SEPEQ-----ETDPAALBEEINH 368
Oy      121 QT-----PMTSTDLTKONFTVDKVNENNAPTYINTSPKSIKKATPKASPKKAAFTV 174
Db      380 TTADDEPAEVAETNDL--EODVNRKVENSEPESEVHDVADATDPVVEAQATETSTNGKFEFNE 447
Oy      175 T-----NPE--IHYYPDNRVEEEDQSQKRDVSVEPLIQOMKMDPQGFNSDSDDTN 223
Db      448 SEFVLSIDONPNQTPTHHEEDAAAEPTVDSTGSESTAEVTAIE-----STVELETA 500
Oy      224 ASV--PPT-----PLHTT-----KPTFAO--LLKKNNVENNSEPE 254
Db      501 AEINNPATFEVEEYLOPRTTITYVDKLDDETYAKPPTVSSSENSVAAEPE 547

```

RESULT 13	TOP1_ARATH	STANDARD:	PRF:	916 AA.
ID	TOP1_ARATH			
AC	P30181.			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	DNA topoisomerase I (EC 5.99.1.2).			
GN	TOP1 OR AR5655300 OR MTE17.1.			
OS	Arabidopsis thaliana (mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
RA	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RA	Kieber J.K., Signer E.R.;			
RL	Submitted (SEP-1990) to the EMBL/GenBank/DBJ databases.			
RP	121			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RA	MEDLINE=990874889; PubMed=9872454;			
RA	Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,			
RA	Tabata S.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. VII.			
RT	Sequence features of the regions of 1,013,767 bp covered by sixteen			
RT	physically assigned P1 and TAC clones."			
RL	DNA Res. 5:297-308(1998).			
CC	-1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE			
CC	CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.			
CC	-1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded			
CC	DNA, followed by passage and rejoining			
CC	-1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH			
CC	NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES			
CC	RELAX ONLY NEGATIVE SUPERCOILS.			
CC	-1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA			
CC	BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN			
CC	WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS			
CC	AT ONE END OF THE ENZYME-SEVERED DNA STRAND.			
CC	-1- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on ways			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).			
CC	-----			
DR	EMBL; X57544; CAA40763.1; -			
DR	EMBL; AB015479; BAB08547.1; -			
DR	PIR; S22864; S22864.			
DR	HSSP; P11387; 1A35.			
DR	InterPro: IPR001631; Topoismerse_1.			
DR	Pfam; PF01028; Topoisomerase_I; 1.			
DR	PRINTS; PF02919; Topoisomer_1_N. 1.			
DR	ProTams; PRO0416; EUTPISMRASEI.			

DR SMART; SM00435; TOPEUC; 1.
 DR PROSITE; PS00176; TOPOISOMERASE I EUK; 1.
 KW Isomerase; Topoisomerase; DNA-binding.
 FT ACT_SITE 872 872 DNA CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 916 AA; 102798 MW; 84BF47913F14454F CRC64;
 Query Match 8.7%; Score 120.5; DB 1; Length 916;
 Best Match Similarity 23.5%; Pred. No. 1.8;
 Matches 72; Conservative 42; Mismatches 111; Indels 81; Gaps 13;
 QY 3 STPSKLLPIDKHSHLQLOPOSSASINSPKPLNPPRTNSKPSLDPNSSDTYTSDDQ 62
 DB 67 TTPSNKTSIYKSS---MPSSSKA---SPAKS---PLRNMP---TYKDRSOLQDKQ 112
 QY 63 EKGEKKKDAFAFQSPDRNDLDSIDIOQTIOHQOQOPOOQO-----LSQDNNMID 116
 DB 113 SECKIEHD-----SEDRPLSSILSGNKGPTSSKQVSSPOEKKNNGDRPLDRASRIIKD 168
 QY 117 EFSFQPMTS-----TLDLTKONPTVDKVENNHA-----PT 147
 DB 169 ESDDEPTISSMFRKKIDSGMSGNQLNDEKKPLVQKHLQNGSTVKNNEVPNGKVLGKRPL 228
 QY 148 YINTSPKSKIMKATPKAPSPKKAFTVYNPEIHYPDNRY-----EEDDSQ 195
 DB 229 EKSSADQSSSLKKAKISAPTSVMKODSVKKEIDKGRVLSFKMAKQSLSTREDGTD 288
 QY 196 KEDSEVPLQHOQKDSQFNYSDEDTNNAVPTPTLHTT-----KPTFAQLNKNNEVNS 251
 DB 289 DDDDDVP--ISKRFKSDS---SNSNTSSAKPKAVKLNTSSAAKP-----KARNVSP 336
 QY 252 EPEALIT 257
 DB 337 RSRAMT 342
 RESULT 14
 HMM3_MYCGE STANDARD; PRT; 599 AA.
 ID HMM3_MYCGE 057081; 049337; 049191; 049370;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytoplasmic high molecular weight protein 3 (Cytoplasmic accessory protein 3) (Accessory adhesin protein 3) (P69).
 GN HMM3 OR MG317.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 [1]
 NC SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE-96011386; PubMed-7592348;
 RA Reddy S.B., Rasmussen W.G., Baseman J.B.;
 RT "Molecular cloning and characterization of an adherence-related operon of Mycoplasma genitalium";
 RL J. Bacteriol. 177:5943-5951(1995).
 [2]
 NC SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE-96026346; PubMed-7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Usterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lueker T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 [3]
 NC SEQUENCE FROM 1-24; 57-169 AND 444-514 FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE-94075230; PubMed-8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;

RT "A survey of the Mycoplasma genitalium genome by using random sequencing";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMM PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE.
 CC ESSENTIAL FOR SUCCESSFUL SURFACE PARASITISM (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT MEMBRANE (BY SIMILARITY).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U39712; AAC71539.1; -;
 DR EMBL: LA3097; AAA99946.1; -;
 DR EMBL: U01716; AAC43190.1; ALT_INIT.
 DR EMBL: U02224; AAA03378.1; -;
 DR EMBL: U02267; AAD12533.1; -;
 DR TIGR: MG317; -;
 KW Cytoplasmic; structural protein; Complete proteome.
 SQ SEQUENCE 599 AA; 68720 MW; D76BE7BD491129A CRC64;
 Query Match 8.6%; Score 119.5; DB 1; Length 599;
 Best Match Similarity 26.1%; Pred. No. 1.3;
 Matches 62; Conservative 36; Mismatches 77; Indels 63; Gaps 17;
 QY 39 PRTNKPESLD---PNSSDTYTSDDQEKKEKKDFAFQTSFDN-EDLNSIDIOQTI 94
 DB 140 PTQQDPSIDAGLPKIEVD-----DOPKAOH---TTELESEPDVELNDSL----- 185
 QY 95 QHQOQPOOQOQLOLSDTNLIDEF-SFQPMPTSL--DLTKONPT-VKVENNHAFTYIN 150
 DB 186 -----PQOQ-----PTENLGDQDFVEKEVPRPQQLHQDLVHQPPQVQSGSNHS----FN 232
 QY 151 TSPNKSIMKKAATPKAPSPKKAFTVYNPEIHYPDNRYVEEDDSQOKEDSEVP-PLIQHQM 209
 DB 233 NSPS---LKPPLVAKPAK---LVQPEVKHPR--QVEQPKPQIVPEKIEPKPEVKH- 281
 QY 210 KDPSPFNYSDEDTNNAVPTP---PLHTTKRFAQLLKN-NEVNSEPALDMLKR 263
 DB 282 -----SHVEIQPKPEVKPVDSVPEVKOPKPVHVEVOPKPVVDLKPQR 327
 RESULT 15
 SCD5_YEAST STANDARD; PRT; 872 AA.
 ID SCD5_YEAST 334758;
 AC P34758;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE SCD5 protein (FTB1 protein).
 GN SCD5 OR FTB1 OR YOR329C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 [1]
 NC SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX PubMed-8688556;
 RA Nelson K.K., Holmer M., Lemmon S.K.;
 RT "SCD5, a suppressor of clathrin deficiency, encodes a novel protein with a late secretory function in yeast";
 RL Mol. Biol. Cell 7:245-260(1996).
 [2]
 NC SEQUENCE FROM N.A.


```

QY 82 ----FELDMSIDIOOTIOMHMOOPMOOOLST-DNNLIDEFSQTPMTSTLDTLTK- 133
Db 2504 SOLPFYSSDLPCDFMOPGLGLOOSPOMHOMGVLOOONIOGOSTINSPTOTFMOTNER 2563
QY 134 ----NPTVDKVNENH----APVYINTSPNKSIMKKAIPKASPKVAFY 173
Db 2564 QVGPFPFVDPSPSIPVGPSPFFSSVKQGHGHLSTSTSQOSPVPSPFTPLPAPF---- 2617
QY 174 VTNPET-----HHYPDNR-----VEE-----DOSQOKEDSVPEPLIQ 206
Db 2618 VANSLSLPGODSTITGHGHSYPGSTOSLIOLYSIDIPEEKGGKKRTRKKKRDADAE----- 2672
QY 207 HQMDPEQGFVNSDEDPNNAVPTPLHLTKRFFA-----QLLNKNNEVSEPP 253
Db 2673 -STKAP-----STPHSDIAPPPPGISETTSPTPAVSTPSELPLQQAQDESVEP 2718

```

```

RESULT 5
US-09-801-368-428
: Sequence 428, Application US/09801368
: Patent No. US20020128250A1
: GENERAL INFORMATION:
: APPLICANT: Busby, Robert
: APPLICANT: Cali, Brian
: APPLICANT: Hecht, Peter
: APPLICANT: Holtzman, Doug
: APPLICANT: Madden, Kevin
: APPLICANT: Maxon, Mary
: APPLICANT: Milne, Todd
: APPLICANT: No. US20020128250A1man, Thea
: APPLICANT: Royer, John
: APPLICANT: Salama, Sofie
: APPLICANT: Sherman, Amir
: APPLICANT: Silva, Jeff
: APPLICANT: Summers, Eric
: TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
: FILE REFERENCE: 109272.147
: CURRENT APPLICATION NUMBER: US/09/801,368
: PRIORITY FILING DATE: 2001-03-07
: PRIOR APPLICATION NUMBER: US 09/487,558
: PRIOR FILING DATE: 2000-01-19
: PRIOR APPLICATION NUMBER: US 60/160,587
: PRIOR FILING DATE: 1999-10-20
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 428
: LENGTH: 605
: TYPE: prt
: ORGANISM: Saccharomyces cerevisiae
US-09-801-368-428

```

Query Match 8.7% Score 121 DB 10 Length 605;
Best Local Similarity 22.9% Pred. No. 0.31;
Matches 65; Conservative 32; Mismatches 65; Indels 102; Gaps 10.

```
OY      21 POSSASAFNSPTKELNPRNTSKPSLDPNSSDDVYTSEODEKGKEKKDPAFQTSFDR 80  
       ||||| : | : | : | : | : | : | : | : | : | : | : | : |  
Db      268 PTSSAPRTSSMTPTSTFTTTTSPSTAASSPTVTYTS-----TTASP 310  
  
OY      81 NFEDLNSIDDIQOTIGHOOQQPQQOQQQLSOTDNNLIDERSFCQTPMSTLDLTONPLVDKY 140  
       ||||| : | : | : | : | : | : | : | : | : | : | : | : |  
Db      311 ITSTITSYNLOTSLEY-----SVLTIVSVHMDPNINISEITSRYLTMKKVY 354  
  
OY      141 -----NENHAPTYINTSPNKSIMKAATPKAPSPKV 170  
Db      335 ITQIISSTLGATPTSASAVATTASVCGRITNNNNNSTNTSNSTPNKSTERKKY-WDSPGKI 413  
  
OY      171 AFT-----VTNPGLIHAYPDNRVEEEDSOQOKEDSVEEPFLIOHQMKDPSQFNV 217  
       || | : | : | : | : | : | : | : | : | : | : | : |  
Db      414 AATFEVVGVCVLVICILIDLILHH--RTRARRAKADPEENEYQSFKYSK-----Y 462  
  
OY      218 SDEDTNASVPPTPLHTTKPTFAOLLNKNNEVNSEDEAL-TDMK 260  
       : | : | : | : | : | : | : | : | : | : | : | : |
```

Db 463 PNEVT-----TTLHTPPS-----SNSTFSTPRLIYTDK 493

RESULT 6
US-09-815-242-11497
; Sequence 11497, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION

```

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US-09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11497
LENGTH: 844
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-815-242-11497

```

Query Match	8.7%;	Score 120;	DB 10;	Length 844;
Post Local Similarity	30.7%;	Prod No 0.53;		

Matches 51; Conservative 42; Mismatches 117; Indels 36; Gaps 8;

Qy 28 IFNSPTKPLNFPRNTSKPSL-----DPNSS-----SDTYTSEDQEKCKEEK 69
: : : : : : : : : : : : : :
Db 134 LFKLPKSVYPPYMNTQSLLKEIYKCOLQAFSPNFSLKGEFENTPSSDQKKETINDKE 193

Qy 70 KDTAFOTSEPRNFELDNS---IDIQOTTQHQQQPQQQQQLSQTDDNNLIDEFSFQTPMTS 126
| : | : | : | : | :
Db 194 KENLKFNPIDENHNTPNEESFLAIPTPYNTTLLNNSSEPDELVOISPHRPHTHTTYLPKRNK 253

```
QY 127 TLDTLKQNPVVDKVNENHAPVYINTSPNKSIMMKATPKASPKKVAFTVNDIEIHHPDNR 1866
    111 111: : : : 11 11 : : : 11 : :
Db 254 FDDILT-NPLLKEPKQ-----ETKEKEPPTLKETPPLILKPIMPISASNTENHDKTEN- 3033
```

```
QY 187 VEEDQSQOKEDSVPEPLIQHQKDPQSFNYSDEDTNASVPPTPLHTT---KPTFAQLL 2433
      : ||| :: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 -HKTPNHPRIKEDDQSPQENQKQENIEENIEKEETO-NAPSFSPLLTTSAKKPVMKEL 3611
```

QY	244	NKNEV	249
		:: :	
Db	362	SENKEI	367

RESULT 7
US-09-765-272-218
; Sequence 218, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:

FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 133
US-09-893-519A-60

Query Match 8.0%; Score 111; DB 9; Length 600;
Best Local Similarity 20.2%; Pred. No. 1.5;
Matches 49; Conservative 37; Mismatches 82; Indels 74; Gaps 9;

QY 15 SHLOLOPOSSASIFNSTK-----PLNFRPNKSKSLDP-----NSSPTTYSSEOQ 62
DB 2 SHSNALPNSP-----FRSPKQRMVEYIGPLNARSFSPVKTPPHGRAGLSPEKRLVLDL 57
QY 63 EKGEKKEDTAFQTSFDFNFDLNSIDIQ-----QTIOHQOQPOQOQOQLOQDNNLID 116
DB 58 DKARRRANSLXNRLMDYLDLTDYLDLBDQRLADRIITKOSRGP-----D 103
QY 117 EFSFQTPMTSLDLTKQNPVDKYNENHAPTYINTSPKSIKKATPKRASPRAVATYN 176
DB 104 EVNGSDVELEIDLTOQRRT-----RRREKKVYV--SS 134
QY 177 PEIHHYPRNRYEEDSOQKEDSVPELLOHQMKPDSQFNYSDEDTNASVPTPLHTTK 236
DB 135 DSSMEYEDTGMPRESSSEEBEADDDGNEVEYVGP-----KERKTSLS SSP--TVK 185
QY 237 PT 238
DB 186 PT 187

RESULT 10
US-09-801-574-57
Sequence 57, Application US/09801574
Patent No. US20020081592A1
GENERAL INFORMATION:
APPLICANT: Wang, Peijiang Jeremy
APPLICANT: Page, David C.
TITLE OF INVENTION: Reproduction-Specific Genes
FILE REFERENCE: 0399, 2007-002
CURRENT APPLICATION NUMBER: US/09/801,574
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: 60/187,518
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: 60/261,557
PRIOR FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 90
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 57
LENGTH: 2789
TYPE: PRT
ORGANISM: Homo sapiens
US-09-801-574-57

Query Match 8.0%; Score 110.5; DB 10; Length 2789;
Best Local Similarity 20.9%; Pred. No. 10;
Matches 71; Conservative 44; Mismatches 104; Indels 121; Gaps 14;

QY 3 STPKSLPIDKHSHLOLOPOSSASIFNSTKPLNFRPNKSKSLDPNSSPTTYSSEOQ 62
DB 2393 SLPSLSLE-----NPKDTCAS-----KSESKIDLTVYSSSHFSGOOEN 2431
QY 63 -----EKGEKKEDTAF-----QTSFDFNFDLNSIDIQOTIOHQOQPOQO 104
DB 2432 LNSMKRRVNVSAETKSDKDKCAFAICDOKSVHGTSPPDHGTLQKFLKN-SPDPYOK 2490
QY 105 QOLS-----QTDNNLIDE-----FSFQTPMTSLDLTKQNPV-----DKVNNHAP 146
DB 2491 SCLSDINBETDVSILVPRDASVLSKPIFCFVKDYNHPLDEM-----NDYFELQDNDIVSSIKN 2547
QY 147 TYINTSPKSIKKATPKRASPRAVATYNPEIHHYPR-----NR 186
DB 2548 SSCMTSPICIONKPIPLQINKLPOTETSEDKYMKDILNPNVTHTGASCHITLNVNO 2607

QY 187 VEEEDSOQKEDSVPELLOH-----QMKDPSQ-----FNYS 218
DB 2608 GAEYLSLEQONKNSKVLQMNATYNNELPQACNPNYSSEHLGTSTPYSAMCYQYS 2667
QY 219 DEDTNA-----SVPTTP-LHTTKPTFAOLLNKN 246
DB 2668 NSGNNAITQTYGITSYEVQSPSGILTTVASTAQOETHSN 2707

RESULT 11
US-09-839-996-6
Sequence 6, Application US/09839996
Publication No. US2003009010A1
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/839,996
FILING DATE: 20-Apr-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RET/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-839-996-6

Query Match 7.9%; Score 109.5; DB 9; Length 1848;
Best Local Similarity 22.8%; Pred. No. 7.5;
Matches 57; Conservative 34; Mismatches 120; Indels 39; Gaps 11;

QY 18 QLOPOSSASIFNSTKPLNFRPNKSKSLDPNSSDTTYSSEOQGEKKEDTAFQTS 77
DB 1347 QAOPTQSTAV---PT--TGETAANSKPAKPAQOAKKPTGEPARENVSTVNTKEPQSOTS 1401
QY 78 FPRNFDLNSIDIOOTIOHQOQPOQOQOQLOQ--TDNNLIDERSFQTPMTSLDLTKQNP 136
DB 1402 -----ATVSTEQPAKETSSVNEQPAPENSINTS-ATVMTETAE-KSDKPO 1445
QY 137 VDKVNNHAPTYINTSPKSIKKATPKRASPRAVATYNPEIHHYPRDNRVEE---EDQ 192
DB 1446 METVTENDQRPANVYADNSVANNSSSEKSRRRRSVSQPK-----ETSAETTYASTQ 1500
QY 193 SOQKEDSVPELLOHQMKDPSQFNYSDEDTNASVPTPLHTTKPTFAOLLNKNNEVNSE 252
DB 1501 ETTVNSVSTP-----KRSRRTRSVOTNSYEPVLEPTENAE--NAENVQGNVNVANSQ 1553

ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 7.8%; Score 108; DB 10; Length 2478;
Best Local Similarity 22.0%; Pred. No. 13;
Matches 60; Conservative 34; Mismatches 95; Indels 84; Gaps 13;

QY 31 SPTKPLNPR-----TNSKPSLDPNSSDFTYTSODODKGEKKDFAFGTS 77
DB 2112 SPTTPNNSDAATGETTATSATDANDKPOANNSSVDASTNSPTMDN----- 2158
QY 78 FDRNFDLNSIDIOQTIOHQOQOPOQOQOLSQTDNNLIDEFSGQTMTSLDLTKONPTV 137
DB 2159 -----DVTSKPEVESTNNGTDDKP-----VTETDN-----ATPAEST---TNNNSTT 2197
QY 138 DKVENHAPTYINTSPNKSIMKATPKASPKKVAFTVTNPELHHYPDN--RVEEDOSQOK 196
DB 2198 TATNEN-APTGSTATAPTASTEAASSADSK-----DNASVNDKONAEVNNASBSOST 2250
QY 197 EDSVEPPLIOHOKDPSOFNYS-----EDTNASVP-----PTPLHTTKPTFAOLL 243
DB 2251 NDKVAQPKSENKAK--AEKDGSDSTNOSWVESTTETLPSADITEPVPNSITSKDEESTT 2308
QY 244 NK-----NNEVNSBP-EALTDMLK 262
DB 2309 NOTDAGOLKSETNVASNEADKSPSKADTEVS NK 2341

RESULT 15
US-09-815-242-12967

Sequence 12967, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12967
LENGTH: 2478
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match 7.8%; Score 108; DB 10; Length 2478;
Best Local Similarity 22.0%; Pred. No. 13;
Matches 60; Conservative 34; Mismatches 95; Indels 84; Gaps 13;

QY 31 SPTKPLNPR-----TNSKPSLDPNSSDFTYTSODODKGEKKDFAFGTS 77

DB 2112 SPTTPNNSDAATGETTATSATDANDKPOANNSSVDASTNSPTMDN----- 2158
QY 78 FDRNFDLNSIDIOQTIOHQOQOPOQOQOLSQTDNNLIDEFSGQTMTSLDLTKONPTV 137
DB 2159 -----DVTSKPEVESTNNGTDDKP-----VTETDN-----ATPAEST---TNNNSTT 2197
QY 138 DKVENHAPTYINTSPNKSIMKATPKASPKKVAFTVTNPELHHYPDN--RVEEDOSQOK 196
DB 2198 TATNEN-APTGSTATAPTASTEAASSADSK-----DNASVNDKONAEVNNASBSOST 2250
QY 197 EDSVEPPLIOHOKDPSOFNYS-----EDTNASVP-----PTPLHTTKPTFAOLL 243
DB 2251 NDKVAQPKSENKAK--AEKDGSDSTNOSWVESTTETLPSADITEPVPNSITSKDEESTT 2308
QY 244 NK-----NNEVNSBP-EALTDMLK 262
DB 2309 NOTDAGOLKSETNVASNEADKSPSKADTEVS NK 2341

Search completed: March 17, 2003, 12:34:28
Job time: 14.412 secs

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:20:59 ; Search time 5.1863 Seconds
(without alignments)
4875.027 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263
Perfect score: 1386
Sequence: 1 MNSTPSKLPIDKSHSLQLO.....NKNNEVNSEPALTDMKLKR 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386	100.0	1664	2 T18216	Integrin-like prot
2	141.5	10.2	236	2 B90577	lipoprotein b (imp
3	138.5	10.0	1166	2 T28680	fibronogen-binding
4	137	9.9	947	2 T08605	hypothetical prote
5	134	9.7	1141	2 E89824	hypothetical prote
6	132	9.5	499	2 S46660	weir protein - pen
7	130.5	9.4	1380	2 S57150	ZMS1 protein - yea
8	130	9.4	1087	2 S58147	protein kinase - f
9	128.5	9.3	1159	2 S62562	probable nuclear p
10	127.5	9.2	1335	2 T30211	autolysin E - Stap
11	127	9.2	736	2 T41259	hypothetical prote
12	127	9.2	1338	2 T18416	hypothetical prote
13	125	9.0	309	2 S57102	hypothetical prote
14	125	9.0	646	2 D82493	conserved hypothet
15	124.5	9.0	773	2 F90537	lipoprotein (impor
16	124.5	9.0	3848	2 T17414	tipc protein - sil
17	123.5	8.9	1094	2 S49313	protein kinase - s
18	123.5	8.9	1254	2 T18277	kinesin heavy chai
19	123	8.9	619	2 D90072	hypothetical prote
20	123	8.9	1139	2 E64234	cytadherence-acces
21	123	8.9	1616	2 G64242	cytadherence-acces
22	123	8.9	2541	2 T29340	hypothetical prote
23	121.5	8.8	529	2 T50609	hypothetical prote
24	121.5	8.8	948	2 T24445	hypothetical prote
25	121.5	8.8	6642	2 T29757	protein UNC-49 - C
26	121	8.7	368	2 G88636	protein W09612.7 l
27	121	8.7	522	2 T40520	hypothetical prote
28	121	8.7	605	2 S48940	hypothetical prote
29	120.5	8.7	910	1 S73361	dnad homolog prote

30	120.5	8.7	916	2 S22864	DNA topoisomerase
31	120	8.7	844	2 B71944	septum formation p
32	120	8.7	1015	2 T13062	CLOCK protein - fr
33	120	8.7	1457	2 T14577	protein kinase Yak
34	119.5	8.6	599	1 A64235	cytadherence-acces
35	119.5	8.6	872	2 S62061	SCDs protein - yea
36	119	8.6	1186	2 S72229	meiotic recombinat
37	118.5	8.5	1390	2 T14004	trifa protein - sil
38	118	8.5	411	2 T40538	hypothetical prote
39	118	8.5	462	1 A60746	chromogranin A pre
40	118	8.5	842	2 S60402	protein kinase CLA
41	118	8.5	879	2 S64755	hypothetical prote
42	118	8.5	1089	2 T14576	nosa protein - sil
43	117.5	8.5	1274	2 A89959	hypothetical prote
44	117	8.4	1062	2 G86325	hypothetical prote
45	116.5	8.4	630	2 T38023	probable transcript

ALIGNMENTS

```

RESULT 1
T18216
Integrin-like protein alpha chain - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C:Accession: T18216
R:Gale, C.; Finkel, D.; Tao, N.; Meinke, M.; McClellan, M.; Olson, J.; Kendrick, K.;
Proc. Natl. Acad. Sci. U.S.A. 93, 357-361, 1996
A:Title: Cloning and expression of a gene encoding an integrin-like protein in Candida
A:Reference number: Z06510; M0ID:96133936; PMID:8552638.
A:Accession: T18216
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <GAL>
A:Cross-references: EMBL:U35070; NID:g1144530; PID:g1144531; PIDN:AAA96019.1
C:Genetics:
A:Gene: alpha INT1

Query Match          100.0%; Score 1386; DB 2; Length 1664;
Best Local Similarity 100.0%; Pred. No. 1.1e-81;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSTPSKLPIDKSHSLQLOPOSSASIFNSPKPLNFPTNSKPSLDPSSTDTYSEQ 60
DB 1 MNSTPSKLPIDKSHSLQLOPOSSASIFNSPKPLNFPTNSKPSLDPSSTDTYSEQ 60
QY 61 DQEKGEKKKDTAFQTSFDRNFDLNSIDIQOTIQHQOQPOQOQOQLSQTNNLIDFSF 120
DB 61 DQEKGEKKKDTAFQTSFDRNFDLNSIDIQOTIQHQOQPOQOQOQLSQTNNLIDFSF 120
QY 121 QPTMTSLDLTKONPVDKYNENHAPTYINTSPKSKIMKATPKASPKKAAFTVNTPEIH 180
DB 121 QPTMTSLDLTKONPVDKYNENHAPTYINTSPKSKIMKATPKASPKKAAFTVNTPEIH 180
QY 181 HYDNRKEEEDQOQKEDSVPEPLIOHQMDPSOFNVSDDTNAASVPTPLTKRTPFA 240
DB 181 HYDNRKEEEDQOQKEDSVPEPLIOHQMDPSOFNVSDDTNAASVPTPLTKRTPFA 240
QY 241 QLLNKNNEVNSEPALTDMKLKR 263
DB 241 QLLNKNNEVNSEPALTDMKLKR 263

RESULT 2
B90577
lipoprotein b (Imported) - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: B90577
R:Chamblaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p

```



```

OY 4 TPSTKLLPIDKSHSLDLOLOQSSASIFN-SPTKPLNFPRTNSK----PSLDPNSS----- 53
Db 303 TSSPLHSTTAHPREKPTSSRAASLSTLSDSEKMTPTSTAAGSPQSPASASTSPYA 362
OY 54 ----DYTSEDQ-----EKGKEKKD-TAFQTSFDRNEDLNSIDIQ----- 91
Db 363 RCGITTSRRRHQINPSSSEKSEPEKKEPSAFET-----LEKSSNVQYKPKSLMPEFLE 415
OY 92 QTIQHQQQQPQQQ-QQLSOTDNNLIDE-----FSFQTPMTSTLDLTKQNPYDKVEN 143
Db 416 KASTGCSFAKQKEGQTSISEKTALSEPEKNKTPVFSFAPASATY--DKSPSPVSISSF 472
OY 144 HAPYIINSPKNSIMKKAT---PKASPKKVAATVYNPPELHHHPDNR-----VEEDQSQ 195
Db 473 NAPSASATKPSPAVSTFSFNAPTTPTSATFSIINKEKPARSPNETIDVDLEEGSGIS 532
OY 196 KEDSVPEPLIQHQMKDPSQFNYSDEDTNASVPPPTPLHTTKPTF---AQLNKNNEVNSE 252
Db 533 AEVEVA-----NEGEDLQKNATIEVAKASISEKVFYFPEAVTDEKNSSEVSS 577
OY 253 PEA 255
Db 578 NQA 580

RESULT 12
T18416
hypotheical protein C0105W - malaria parasite (Plasmodium falciparum)
C.Species: Plasmodium falciparum
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C.Accession: T18416
R.Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, July 1997
A.Reference number: Z18934
A.Accession: T18416
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1338 <LAW>
A.Cross-references: EMBL:Z97348; NID:e1323671; PID:e1323672; PIDN:CAB10568.1
C.Genetics:
A:Map position: 3
A.Note: C0105W

Query Match 9.2%; Score 127; DB 2; Length 1338;
Best Local Similarity 19.6%; Pred. No. 1.5;
Matches 70; Conservative 50; Mismatches 109; Indels 128; Gaps 15;

OY 16 HQLDQPOSSAIFNSPTKPLNFPRTNSKPSLDPNSSSDTYTSSEDOQEKGE----- 67
Db 207 HSDIKPE---NVLVSPLTTIPKPKDYTKDKLESKNSNOVEKKENDQNVDKLITTMNN 262
OY 68 -----EKKDTAFQF-SFDRNFDD-----NSIDIQOTIQHQQQPQQQQQLSQ--TD 111
Db 263 INTNLSEKKKVVINDQKNDKNIETQCKTSSKENIEDNVSEFVNDPSDPQKNNLNNTD 322
OY 112 NNILIDFESQFQTPMTSTLDLTKQNPYDKVENHAPF----- 147
Db 323 NNIIIS-NVQIEKQSLSKNKKNEKQSYININNSLINDQNLKREDIKFNDAEGITKYD 381
OY 148 YINTSPKNSIMKKATPKKSPKK-----VAFTV-----TNPELIHHYPDNVEE- 189
Db 382 MLNIKNNISIKERINDCHSPNEWKNDNNOCEDNSINICNNKNNNIQTNNIINDNTVNEK 441
OY 190 -----ED-----OSQKEDSEVPEPLIQHQK 210
Db 442 INNTSKDMLNNTQNNNDSEKNDVYIEQQLVNEIDLKKNKKQYKPKKNNIPEPPYVNHKR 501
OY 211 ----DSQF-NYSEDEDTNASVPPPTPLHTTKPTFAQL-LNKNEVNSEPEALTDML 261
Db 502 PMSNDPSLLTSTSN-----IHALQETLTKPKPYHHYNTPLNPEKTRDKNM 546

RESULT 13

```

[illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:19:45 ; Search time 58.7193 Seconds
(without alignments)
5839.012 Million cell updates/sec

Title: US-09-964-858-1

Perfect score: 8631

Sequence: 1 MNSPTSKLPLDKHSHLQLQ.....WVNLMLQQQQQQQQSSSQ 1664

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

otal number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_bacteriap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	382.5	4.4	3257	5 Q9V736	Q9V736 drosophila
2	381	4.4	2081	10 Q9LH98	Q9LH98 arabidopsis
3	369.5	4.3	2647	5 Q9U4X0	Q9U4X0 plasmodium
4	369	4.3	1723	5 Q8T224	Q8T224 dictyostell
5	367	4.3	704	3 Q9P7Y8	Q9P7Y8 schizosacch
6	355.5	4.1	1514	5 Q8SY55	Q8SY55 drosophila
7	355.5	4.1	2910	10 Q9FND5	Q9FND5 arabidopsis
8	346.5	4.0	2771	5 Q26216	Q26216 plasmodium
9	338.5	3.9	5327	5 Q76891	Q76891 drosophila
10	337.5	3.9	1738	5 Q76329	Q76329 dictyostell
11	337.5	3.9	16215	5 Q9NFS3	Q9NFS3 drosophila
12	336	3.9	6815	5 Q9I7U4	Q9I7U4 drosophila
13	335	3.9	2678	5 Q9NDS4	Q9NDS4 dictyostell
14	333.5	3.9	3111	5 Q9VH10	Q9VH10 drosophila
15	331.5	3.8	3147	5 Q17464	Q17464 caenorhabd
16	331	3.8	2472	5 Q8T2M5	Q8T2M5 dictyostell

17	329	3.8	6713	16 Q99U54	Q99U54 staphylococ
18	329	3.8	1807	5 Q931R6	Q931R6 staphylococ
19	327.5	3.8	1607	5 Q8MSL1	Q8MSL1 plasmodium
20	325.5	3.8	2768	5 Q9VC00	Q9VC00 drosophila
21	325	3.8	1650	5 Q77328	Q77328 plasmodium
22	324	3.8	1302	2 Q49547	Q49547 mycoplasma
23	322.5	3.7	3354	5 Q8T101	Q8T101 bombyx mori
24	320	3.7	5085	11 Q9JKS6	Q9JKS6 rattus norv
25	320	3.7	17352	5 Q95YM2	Q95YM2 procarnarus
26	318	3.7	1236	5 Q9GTX2	Q9GTX2 plasmodium
27	317.5	3.7	3130	5 Q9BK46	Q9BK46 plasmodium
28	316.5	3.7	1593	5 Q8T1T7	Q8T1T7 dictyostell
29	316.5	3.7	3254	5 Q9BK45	Q9BK45 plasmodium
30	315.5	3.7	1616	4 Q96PH3	Q96PH3 homo sapien
31	314.5	3.6	2867	5 Q9N2M3	Q9N2M3 plasmodium
32	314.5	3.6	3724	5 Q77320	Q77320 plasmodium
33	314	3.6	4498	5 Q9W223	Q9W223 drosophila
34	313.5	3.6	3484	5 P91257	P91257 caenorhabd
35	312.5	3.6	1769	5 Q16625	Q16625 caenorhabd
36	312	3.6	2038	5 Q967Y0	Q967Y0 dictyostell
37	311.5	3.6	1803	5 Q9VU16	Q9VU16 drosophila
38	310	3.6	1492	5 Q8SSU1	Q8SSU1 dictyostell
39	309.5	3.6	1271	5 Q25860	Q25860 plasmodium
40	308.5	3.6	5147	4 Q9Y6V0	Q9Y6V0 homo sapien
41	308	3.6	2454	5 Q8T2G3	Q8T2G3 dictyostell
42	306.5	3.6	2062	4 Q9NG231	Q9NG231 homo sapien
43	306	3.5	2151	5 Q9NG79	Q9NG79 trichomonas
44	305.5	3.5	1204	5 Q8T134	Q8T134 dictyostell
45	305.5	3.5	2752	5 Q9BJY0	Q9BJY0 plasmodium

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	3257 AA.
ID Q9V736	Q9V736		
AC Q9V736	Q9V736		
DT 01-MAY-2000 (TREMBLrel. 13, Created)			
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE CG12864 protein.			
GN CG12864.			
OS Drosophila melanogaster (Fruit fly).			
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC Ephydroidea; Drosophilidae; Drosophila.			
OX NCBI_TaxID=7227;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=BERKELEY;			
RX MEDLINE=20196006; PubMed=10731132;			
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,			
RA Brandon R.C., Rogers Y.-H.C., Blaziel R.G., Champe M., Pfeiffer B.D.,			
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,			
RA Abrell J.F., Agdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA Borikva D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,			
RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunlov B.C., Dunn P.,			
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,			
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA Hostin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,			
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,			

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Modarity C., Morris J., Mostreli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Neout D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacel J.M.,
RA Palazolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003814; AAF58230.1;
DR FlyBase: FBgn003946; CG12864.
DR InterPro: IPR000637; AT hook.
DR PRINTS: PR01608; BACINVASINC.
DR SMART: SM00384; AT_hook.1.
SO SEQUENCE 3257 AA; 355970 MW; 7E9212C20A40810C CRC64;
Query Match 4.4%; Score 382.5; DB 5; Length 3257;
Best Local Similarity 19.6%; Pred. No. 3,4e-08;
Matches 373; Conservative 292; Mismatches 679; Indels 559; Gaps 88;
QY 4 TPKSLPLD-----KSHLQLOPSSASIFNSPTKP-----INFRTNSKPSLDPS 51
DB TGSSTLPODDIEMASNHQETDLKCAPDRVALDKSESTPRVEEQICKVPTPDTALDESK 647
QY 52 SSDT---YTSDDQDKGKEE-KKQPAFQTSFPRNFDLNSIDIQOTIQHQOQPOQOQO- 106
DB 648 VSESKKNHIELEDKKDKKEQKESPNKSKETN---ENSIVTVEVELPAKKAKEA 704
QY 107 ---ISQTNHIDERSF-----OTPMSTLDLTQONPVQVNVNHNAPT 147
DB 705 GNIVESSQALAEKLAELIAEYKGEANEVSVTSVQG--EONPIYELVKELEDET 762
QY 148 YIINSP-----NKSIMKAPKSPKVAFTVTNPETIHHPYDNRYEEDOSQOKED 198
DB 763 ISEVVPANODSSVEDQTLADKENVEKPRPVKAPSS-----KDEPPAEENLPAPQD 816
QY 199 SVE---PPLIOHOMKD-----PSQFNYSDD--TNASVPRPPLHTTPAQQLKNNE 248
DB 817 PIEQKTPVAKNQQHDKHNEAPRAESLSVSDIPSSVTPSKRNHSSP--ANTPKKSE 874
QY 249 V-----NSEPALTDMKLRENFNLSL--DEKV-----NLYSPTNNN 286
DB 875 IEALQSSVPRAALSKDKATPQNLRESRSKRLKLTETLLMDTMRSSPRLGRSPAESH 934
QY 287 SKNVSDMS-----HLOLQDASKNK-----TNE 310
DB 935 SHERSPEMKVTVSKLAKDLTITIDKEKEIELKSLPDASEFQVKTITKTASDTSILDE 994
QY 311 NIHLSTALKAPKNDIENPLSLTNADISLRSSGSSOSLSLRNDNVLSVPG--SPK 368
DB 995 NPSSSKTEMKLKG--KPLK-----AKKMSRTSETEVKAIALDSN--EDIPSISFISIK 1042
QY 369 KVNGLSLNDIGKGFSEVEVSLPR--DLSDRLKLETTKHDAPEHNENFNTDKKSTNT 425
DB 1043 CVEHLLSSSEEQ--KDEKELLCPKPOIDCTNIDLE--OSTALETDTEO--VEKKRSNR 1096
QY 426 NKGOLVSS-----DHLDSFDSRYNHTEOSILNLSASOSQISLNALEKOR--- 473
DB 1097 RKSRRINERKFTETDTLHDLDKAKENASLEISMFKCTLETOQSPRYAKKKRNSGR 1156
QY 474 -QOQEOQOTQAAEPEETS---FSDNTKVKQEPKSNLEFVAVTTIKKEVSAETELAKPRE 529
DB 1157 LSRREKSVINAASEKKSPPAISOSTERQQLNEN-----PSKKD--KKTQESGKNKE 1208
QY 530 FSSRLIKNEDEIAPADIIHPKKEEA--NSHVEDTALLKALANDEESTJTQNSTKMS 588

DB 1209 ---AVQPLDKETSSSTNIIDKKNSEFDSAQPSD-----RLNOKESAFTKLSISSP 1260
QY 589 IRFHISDMKL-----EDSNGDRENDNDIRFEKSDI-----LNDV 635
DB 1261 KTKMKQDKDLALSKGSDSPNPTIRPTGEDSRQTDKHQDNQKHEEDSSKLKANIDET 1320
QY 626 SQTSDIIGKRYGNSSEIETTK--TLAPRSPDNDKENSLSLED-----PANNESTL 673
DB 1321 KSSSEKDAEPIKSDSSQDAKAPRLSKPSKNNKKKKKKKKKPNDSIAESDIEGQVQVTEY 1380
QY 674 QOOLEP--HTKEDDSLANSNAP-----PEELTPVPEANDY 711
DB 1381 QATCSPSESNKKDKWKSQDETNEEPLNSETELGRIRKQCAFHIENPKD--DLHITPONEN 1439
QY 712 SSFNDYTKTFDAVSSPEESISREHENDSKPINFISIMHKOEKKKQIKQIKVPTKQIIAS 771
DB 1440 QSIAGV--NFEKQVLPRES---ESDT-----PIKTIPTKTYLMCT 1475
QY 772 QOQKNDQESRVTSDKVKIPNALQFKKKEVNWASRRVSPMDDLNVQFLPELSEDSGF 831
DB 1476 KNTSLSLASEDDIYLEP-----QKLIITSKGDSNPDLDMAN-----NLETSS 1520
QY 832 KDLNFANYNNNTNRPKSFPLSTKVNLSINDDPNVPEP--KSYAEIRNARLSANKA 890
DB 1521 QPKHEHESDQT-----FT--DNSDIIPCTKKSQIVFTPTPKSSDQTKN-----SFI 1567
QY 891 APNOAPLP-----PORQSTRSNSKRRVRRVRYFE-----IRRS 929
DB 1568 TNRNRSKSKRNYSKEAKRLDNSFEESQNAASESSAKVOKELTPTASCRLKRLVLRKP 1627
QY 930 SALAPCDMYNIDFEDGAGSKPTIK--AEGMKTLPSMD-----KDVKRI 972
DB 1628 TSLPTNSRKSIFKKTPAKSKRLTKLEMEKTPSEPSVLSGEVNPDPMAAESVAVL 1687
QY 973 LNAKGYODEYINAKLVQDKKNSIYD-----PEDRYE-ELQO--TASINATIDSSI 1025
DB 1688 HESDHDLESNEIPNEVEVEDTEEASAEQDNKLRKEDDELEVNDICAAKSNPTDOST 1747
QY 1026 YGRPDSISF--DMLPYLSDELKRPITALSAD-----RLMEQEVNHLRS----- 1068
DB 1748 KQASSNKSPTSDSVLDQETDELNS--NSLINATQGEDTPTKELTEEVPRNNKTVEDESKKQ 1804
QY 1069 -----NSVLYHPGAGAAATNS-----SMLEP-----DFELINSPAR 1099
DB 1805 EILKDLPEPDNALIEDTASTAKAAEEMDLYIKESNVKSVLAEPETVDYDDEALQSPRP 1864
QY 1100 NVSNNSDVVAISGNASTISF-----NOLDKPNPDQAT-----IGQKIQEOP 1140
DB 1865 N-SETTIVTDDEPSTSSVVKRSILRKREADSSQDEAKRKQRODVEKSLTGKKEQYKP 1923
QY 1141 A-----SKSANTVKG-----DDGLASAPETPRTPKKSISSK 1174
DB 1924 ARRLQLAEEERPSLKRKTESEAKSTYQKYSIIGNETIMSTTARIPRETNREASTS 1983
QY 1175 PAKLSA-----SPKSPFKIGSFVVRKIKKNGSIAGLETPKAT--HK 1215
DB 1984 PSARKSAVOEAKHVTETTHIILGPPGKLLHSDSPAALVKK---PMVOTLLSTLSLOK 2039
QY 1216 PKKSFQNEISNHKR-----DGGISPSGSEHQQHNSMVSVSQIYDATST 1263
DB 2040 PSTLDDGSPV---KIRKSLKSIADENIDGQOSIFSSSSVVLNKNYSVA--PRKVNISVSL 2095
QY 1264 VPDEKKNQV-----HKPREKOKONNNNNNNNNNNKOKTDIRGVVDDELR 1307
DB 2096 L--QSKDTQVETLAASSSTETPLTKKEKLTQKSTKPRGNNKTESKKSLSVQG-----P 2147
QY 1308 DVLQERKLEFRVLGKININ--LPDINTNKGFTLLDNGVNCVUTPREYNNMDHNVAIGK 1366
DB 2148 QMTQKQSE--AVSGPILKYLKSETPSSRKIVSYVTG-----RKQIQG 2190
QY 1367 EFELTVADSLFILTAKRSYKPRGTLVEUTEKVVVSRNLSRLFGSKDITTTTKVEVPT 1426

Db 1791 GNDNLSLTJSTSE-----PBGKLD-TNKDSMKN-WTMEAGGS 1827

RESULT 3

ID 0904X0 PRELIMINARY: PRT: 2647 AA.

AC 0904X0:

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Putative erythrocyte binding protein EBL-1 (Fragment).

GN EBL-1.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TextID=5833;

RN [1]

RP SEQUENCE FROM N.A.

ME MEDLINE=20078864; PubMed=10613703;

PE Peterson D.S., Wellem's T.E.;

"EBL-1, a putative erythrocyte binding protein of Plasmodium falciparum, maps within a favored linkage group in two genetic crosses."

RL Mol. Biochem. Parasitol. 105:105-113(2000).

DR EMBL, AF131999; AAD3018.1; ..

FT NON-TER 2647 2647

SO SEQUENCE 2647 AA; 304550 MW; AE98F88FD754E300 CRC64;

Query Match 4.3%; Score 369.5; DB 5; Length 2647;

Best Local Similarity 19.5%; Pred. No. 9.5e-08;

Matches 383; Conservative 305; Mismatches 717; Indels 555; Gaps 95;

QY 21 POSSASISFNSTKPLNPTNPKSDS-----LDPNSS-----SDTYSEDOQ--- 62

Db 825 PSSSKAL--KPKIKDVFEIEETKSELSLTDKSKNTNPGSGGNGYDROISKRDVAH 881

QY 63 -----EKKEE--KKDTAFQTSFDRNFDLNSIDG-----QTOHQOQ 99

Db 882 DCPKEKSGEKEVPKIDAAVKT---ENFTSNRNDGEGKEKSGDHSFVHSHDKIKNEP 938

QY 100 QPQOQOQSLQOTNNLIDERSFQTPMTSLDLTKONTVAVKNEHAPTYINTSPKSIK 159

Db 939 QKVESENLPIKEKM--ESSDSIPIT-----HIEAKGQSSNSDMD 978

QY 160 KATPKASPKKAAFTYNTPEIHHYPRDKVEEDOSQOKEDSVPRPIQHQKDS--QFNYS 218

Db 979 PAVVSGRESKDVNLHTSERIKENEKGVITKDDSKSIEIS-----KIPSDQNNIS 1028

QY 219 DDDTNAVSPRTPLHTTKTFAQLKKNNEVNSEPEALDKMLKRENFSLIDE--RV 275

Db 1029 DLSQAN-----EDSNQCKKETINPPS--TEKNLKEIHYTSDSDHGSKI 1072

QY 276 NLYLSPTNNNNSKNVSDMDSHLQNLQDASKNTNENINHLSPALKPKNDIENPLN--SL 333

Db 1073 KSEIEPKELTESPLTDKKTESAIGD---KNHESVKSADIF---QSEIHNSDNRKI 1124

QY 334 TNADISLRSGSSQSSLOSLRDNR-----VLESVPSPKVPNGLSLNDGKIKFSQE 386

Db 1125 VESVYQDSSGSMST--ESIRTDNMDKFTSEDIAPBSINGHEK--IGSSADD--RGSSE 1177

QY 387 VESLPLRDLRDLKLETTKEHDAPEHNNENFIDAKS--TNTNKGGLVSSDHL----- 439

Db 1178 --KSIIDKSEFENKNSHSDIKQSDNCGSTDYELTESPKGLDESVPSSIDMLKP 1235

QY 440 ---SFDRAVNHTEOSTLNLNSASQSLNLAKOKROTQEOEQTOAPPEETSPSDNI 496

Db 1236 NKSPVTSFSDHVDSPNISLQASQN-----ADSYGGEKPSNNILATDGD 1280

QY 497 KKKQPKSLLEVVKYTIKKEPVSAIEIKAPKREFFSRILRIKNEDEIAPADIHPKKE 556

Db 1281 VSEKE-----KISVSPNVSVTYDEGDKRQ-----GISDSSSIHHEIDE 1320

QY 557 ANSHVEDTALLKALNDEESDTQNSTKMSIRPHIDSDWL-----EDSND--GDREBN 610

Db 1321 KNLWE--SFISGGL---EEGDIKEKCK-----EDGSLIPISPEKINDFGKRENI 1367

QY 611 DDISREF-----KSDILNDVS--QTSDIIDGKYG-----NSSSEITTKLAPRSDNDE 659

Db 1368 VDBSVSEYVDNRKSDISNVSEEOSSITISRRGTETGINNSEELTSEFTSVDNVNRDH 1427

QY 660 NKSLEDPANNESLQOOLEVPTKEDDSLANSNNAPEELTLPLVEANDYSSFNVDYK 719

Db 1428 NMO--ENLVSSSTQOESEREKREKENADSHSELSLSIEVGETIRRDADAESEND--K 1483

QY 720 TPDVASSFEESLSREHETDSKP--INFISITHQOKOKIHIKVPKQIILASTYQOYKNEQ 778

Db 1484 GBDIIOSEEOQIVTEKTKLIEPTVNL-----QPSPLIEIHKESSEI-----DDKDMT 1533

QY 779 ESRVTSKDKVIRNALQFKKFEVNVSRVVSPPMDDLNVSOPL--PELESDSGFKDLNPA 837

Db 1534 GGEVHTTDA---NTROYRSESEVEYITNKPDMEKMTTKPPSOITIEKEIETDSTKN---- 1586

QY 838 NYSNNTNRPRESFTPL-----STKNVLSN-----ID-----NDPNV 867

Db 1587 --QDNDEQNSIIPLINKNEEDGVSIPSTRVWESGFSVRNEQIIEKDKHITDDPT 1644

QY 868 VEPREP--KSYAEIRARLRSANKAAPPLPQRO---PSTRSN--SKRVSRFRVP 921

Db 1645 INPSENGLGYEMPN---DSIKSVTITESPLRDVEOMIEPIDGNGKNNIIEGPOS 1700

QY 922 TFEIR-----TSSALAPCDMY-----NDIF---DDFGG 948

Db 1701 TTEIRQOMGPSNVNIPPELIPVAGSKLEAKERSMDADKGTITETIYVEEDPNGIG 1760

QY 949 SKPTIK-----AEGKTLPSMD--KDVYKILMAKGYTODEYINAKLVQDPKKNSTIV 1000

Db 1761 EHONLKEVHEQASELNTYNSLDGRTVVEKERLDEPQSIIPDRITTEHIEIDKEKE--I 1818

QY 1001 TDPEDRYEEL-----QOTSINATIDSSIYGRPOSITDMLPYLSDLKRPPTALLSMD 1055

Db 1819 HEP---NELDHNGEOEEMIRNEVSDNRK---DEQISDKNETROLQDHEDSDODEITD 1871

QY 1056 RLFMEO--EVHPLRSNSVLHP-----GAGAATNSS-----MLPEPDF 1091

Db 1872 KREMENTLEENPNSSSDSLNPEGKEKGTIEHHSSELDVSVDYKYKGIENOITEPI 1931

QY 1092 ELINSPARVNSNNSDVAISGNASTISFQOLDMNFDDQATQOQIOBPASANTVKG 1151

Db 1932 ESVEQODTINPGNSKETEDVN-----SRADMEEKDVKIKIITEE-----S 1973

QY 1152 DGLASAPETPTPTPKESISSKPAKLSSASPRKSPIKI-----GSPVAVI--K 1198

Db 1974 EEELEISKDTTSHSEKPSIEOSVNIUDS---KNEINQIEKNVONQONEDDPILLEE 2030

QY 1199 KNGSIA-----GIEPIK 1211

Db 2031 OKNKIATLEQNKNEYNPNPQLSHKERTLEVDLEQSMOTNCLTSELNKKCDSIQIPE 2090

QY 1212 ATNHKPKKSFQGNELSNHKKRQDGISFSSGSEHQONPNVSPQSYDTATSTVPEBNKY 1271

Db 2091 ASNTDK--LMDKITENKQDFSEIEKSVGEIH--ENKDKLINESAESD--DVPQNK-I 2143

QY 1272 OKHREKQOKNNHNNHNNHNNKOTD---IPGVUDEIDVYGLQERGLFFVYLGKININ 1328

Db 2144 EHDSEAGVTDQYETDYEEKGVNFVGLPKPREBERDV---TRG-----S 2188

QY 1329 LP-DINTHGRFTLLDNGVNCVTPPEYMMDDHNAVIGKFEFLV-----ADSL 1376

Db 2189 VGSVDAPFEKLIHERLDSN-----EKNKEELANS--KEIERYGLDILGIDSESBSDV 2239

QY 1377 EPLITLAKAYEKRGTLVTEKKYVKSNNRSLRFGSDIIL-----TTTKFVPTVEK 1429

Db 2240 TI-----YKRPDTFVENVH--VLSHD--IKSLFENKKTGVLNKOQKETEKNIAGMI 2288

QY 1430 DTWANKFAPDGSFARCYIDL--QOFEDQITGKASQ--FDL-----N 1466


```

OY 1295 KTDJPGVVDDEIPVUGLOERGKLEFY-----LGKINLNPIDITHKGRETL 1341
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1275 SSSLGAGFQFTIQOOOQOOSNIFYNSPYNSSOVYNNPRTGSTITTSLAGPST----- 1327
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 1342 TLNDGVHCVTTPXYNMDDHVAIGKE-----FELTVADSEFLITTLKA 1384
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1328 TSSAMQHMIT---NMTSSNIYVNQQNNNNNDQNNNNNNNNNNSTTNVANNNTTNT 1383
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 1385 SYEKPRGLVEVTKKKYVKSNHLSTRFGSKDILTTTKKYPRPEVKDTMANKAPDGSFAR 1444
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1384 PSSPPQPNCTPTGTSTISNGVVPSLOWRGITM---PREVLPT-----PD----- 1428
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 1445 CYIDLDQEDDITGKASOFDLNCFENEMETMSGNOBPKRGRPKYLQAEVKKLYLPURSP 1504
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1429 -TLVLOOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQDETPTPTPSNIS--PRSP 1480
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 1505 -----RELTPTSIRSAVESINELNECONNYFEGLYOEGBDCPIFRKREFKLWG 1553
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db, 1481 VHOOSPSTNTTTTTSTTTTIR--HSAYTOLS-----FAGLHNQVS--PISPRSPDHG 1530
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
      1554 TS-LLAHSEISHKTRAK-----INLSVUVLIYDKENDRSNHRF-----SDVL 1596
          ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
      1531 TSGDYNDGOSOPSSRRKNRPETFOLIKRMNDCF---ENLDKNNGKFUSEICQATLEG 1586
          ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 1599 LLDNAFIKFNANGELIDFCAP-----NKHEMKIWLIOEIIYNRRPORPW 1645
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1587 LTDOQVRVFQFKNRABSRPSPRGQPTNPLTSSNTGNONSMLAQHLC-QHLOQVOQOQO 1645
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 1646 VNMLAQOQOQOQOQOQOQOQSSQ 1664
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1646 QLLDQLQOQOQOQOQOQOQLHAQ 1664
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

ID	09PY78	PRELIMINARY;	PRT;	704 AA.
AC	09PY78.			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	Hypothetical 78.7 kDa protein.			
GN	SPAYUG7.03C.			
OS	Schizosaccharomyces pombe (fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomyces.			
OX	NCBI_TaxID=4896;			
	[1]			
	SEQUENCE FROM N.A.			
	STRAIN=972H:-			
RA	McDougal1 R.C., Rajandream M.A., Bartell B.G., Saunders D., Harris D.;			
RL	Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AL316521; CAB66312.1; -			
DR	InterPro; IPR001849; PH.			
DR	Pfam; PF00169; PH: 1.			
DR	SMART; SM00233; PH: 1.			
DR	PROSITE; PS50003; PH_DOMAIN: 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 704 AA; 78716 MW; 755E28CD67F127AE CRC64;			
	Query Match	4.38;	Score 367;	DB 3; Length 704;
	Best Local Similarly	21.0%;	Pred. No. 2.4e-08;	
	Matches 189; Conservative 140; Mismatches 315; Indels 256; Gaps			
QY	767 VKIPALIQFKK-----EYWNMSRRVYSPDDDLNVSQFLPELSDESGFKDLNFAYS	840		
DB	21 LRIBSPSPIDYECSDYASTIASISTRESTMRNFRNSISSTAFSESEDAEDGDSFPYD	80		
QY	841 NNTNPRSF-----TPLSTK-----NLSINDIDPNVVEPEPKRSVAETIRNARLSA	887		
DB	81 OTLSNSSSFDDHQSLPFSTEVRRTPPYSVANETDSSSTVE-----DVNKENIISL	132		
QY	888 N-----KAAPNQALPPPPQRPQSPSTRSNSNKKRYSRRVPTFEIRRTSSALACDWDYNDIF	942		

Dd	133	NDSCILIKIDDEA-----		164
Qy	943	DDFAGSGKPTIKAGMKTLPSMDKDDVKRIILAKKGVTODEYINAKLVOKRKNISIVTD		10020
Dd	165	-NOGHGDIPI-----		177
Qy	1003	PEDRYEELQOTASIHNAITIDSSITYGRPDISTIDMLPYLSDELK-KRPIALLSADRLEPMEO		1061
Dd	178	-----PARVSCKSL-----FNEBTDPAEEVEISIPPKYL-----EL		210
Qy	1062	EVHPLRNSVYLHVGACATNTSSMLPEDPFELINSPPARVNSNSNVNVAISGNAPISFENO		1121
Dd	211	PTH-----SHNS-----DPSFNS-----IYSSVSDMGWGBECSINIA----		244
Qy	1122	LDMMFDQATIGOKIOEQAPASKSANTVRCDDGLASAPETPTPT-KKESISSKPAKLSS		1180
Dd	245	-SFGFSEBSSSFODIKTPRLSPADENR-----ENCRTDIYRSDSIHEYEDEPLTS		293
Qy	1181	ASPKSPPLKIGSPRVYIKKKKSGIAGEIPRIKATHKKKKSGFQGNELISNKKVRRGSGISPSG		1240
Dd	294	-----SITSLSDP-HUVDEN--APRLPLPKVYSLP-----		320
Qy	1241	SEHOHNDSWVSPSOYTDTATST-VBENKDVOYKREKOKOHNNHNNHNNHNNKOTDIR		1299
Dd	321	-----DBREFNVLSAPDALTPTTLRKQNSKVUHNATSKOEMQTSRRVUNSCWMPES-LS		373
Qy	1300	GVVDEIPRDVGLDERCKLFFRVGLGGINIMLPROLNHKKGFTLTDGVCVYTPREYMNDD		1359
Dd	374	RNLSSSIQOQTG--GSGKLFYRLMEIRNLTPRLASGMTTFEYTTI--SGKH-IQVP-WNLHLH		428
Qy	1360	HNVAIGKEFELTVADSLFELTLTKASYEKRGRTGLVEVTEKKVAKSNRLSLFGSKDIT		1419
Dd	429	STKIENEYTFDESIISSIVCTLRAAYDPK-----VTRRSITGKVFSF-----N		473
Qy	1420	TTKRYVPE-YKOTPMANKFAPDGSFARCYIUDLOQFEQITGKASQFDLNCFMEN--ETMSN		1476
Dd	474	KRKMTTIDPVSEALHGFVSBDGTGEVYTIINTOSVSTALGRQOSMVLPIMKMTVIDPRAK		533
Qy	1477	GNOPMKRGKRPYKIQOLEVVKMLVYVPSRPREI--LPTISIAVESINELNNEONYFEYTL		1534
Dd	534	DVKPL-----PRKVELEIHFIPLR-ALPYSIKLPRASIESAMYDKLAEIMORTLLCDSYL		588
Qy	1535	HOEGSDCPEFKKREFKLMGTSLASHSEISHKTRAKINNVVDLI---YVDKENIDRSN		1590
Dd	589	COQGSGDCPYWRRRYFOLIGSKLVAFOOFSKVVRATIDLSSEATHIYDDNNYSOELE---		645
Qy	1591	HRNSVDVLLDHAFKIKFANGCELLDSCAPRKNHMKIWIQNOELIYRNFRQRPVNLML		1650
Dd	646	-----GYUFESGRIIFSSNGDYIDYIAETGVGKEDMSWSTLQNHIGQCSWNNKNTKTSFL		700

```

RESULT 6
ID      08SY55      PRELIMINARY;      PRT;      1514 AA.
08SY55
AC      08SY55;
DT      01-JUN-2002 (TREMblref. 21, Created)
DT      01-JUN-2002 (TREMblref. 21, Last sequence update)
DT      01-JUN-2002 (TREMblref. 21, Last annotation update)
DE      GH09355p.
CN      CG6004.
CS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxId=7227;
RN      [1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=BERKELEY.
RC      Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA      Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA      George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA      Miranda A., Mundall C.J., Nunoo J., Paclebb J., Paragas V., Park S.,

```


QY 124 MTSIDLUTK-----QNPVVKVNNENHAPTYINTSPKNSIMKATPKASPKVAVFTVNPET 179
DB 850 V-SLLVKSVEOMOKPPL-----SPSEVSEETSKTVDEKIEEK--PEE-----EV 892
QY 180 HHPRDNREVEDDSOQOEVEPPLIOHQMKDSOFNYSDEDTNAPRPLHTTKPTF 239
DB 893 TLVQEOGVDSYGLTEKKEEVSVP-----ESTELEPOFQOEBSVJIDPTLPLOKPTLES 945
QY 240 AOLLNKN-----EVNSPEALTD-----KIKREN 265
DB 946 SEVLEESSKTVDEKIEKTDSIELGEIAQERSVTDLTPLOEBSQOPNEQETKLEKHE 1005
QY 266 FSN--LSLDEKVN--LYLSPTNNNNKNSVSMDSHLQNLQDASKNTNENTHNLSPALKAP 322
DB 1006 PTNEEKSDEVEVLASPSKELEGETVVE---AENIENIKNEEBOAAEKLOKSLSEV 1061
QY 323 KNDIENPLNSL-----TNADISLRSSGSSOSSLOSRLNDN---RYLESVPGSPKKV 370
DB 1062 QT-VESPSILFSSEBODHTVAEIVYDEKAKEEVPMLQIKNEDDATKIHTREVEQARDI 1120
QY 371 NPGCL-----SLNDGKGFSEVVESSLPRDLSRDKLETKEH-----DAPEH--- 412
DB 1121 GBSLFTGCSIN---QNOPEOVKEACSKBEQKEISTNSBNIVNETYALHSVAAEBETA 1177
QY 413 -NNENFLDAKSTWT-----NKGOLVSSDDHLDSPRSTNHTBOSTLNLNSASOS 462
DB 1178 TNGESLDDVETKSVLLEVRKEEEMKTDAPRIDAIEKELEEVKTVVQADAKTVNNE 1237
QY 463 QISLNALEKORQTOEOEOQOAAPEEET--SFSNDIKVQOPKSNLE---FVKYTIKREPV 518
DB 1238 ETJAHSESLKQGNHOKN--APVEATQNLDDABOISREVTYVDTREADITKIEKVQE 1295
QY 519 SATETAPKRRSSRLRIKINED-----ELAPADIHKKENANSHEVDTD----- 565
DB 1296 GPVVIETP-----TIQEDIESETSLKEEVD--OSSKDEHEHVLERDIPQCE 1345
QY 566 -----ALLKALNDESDDTQNTKMSIRPHIDSMK----- 598
DB 1346 LAEAVDTSTVEBAITLKTLETNISPEAMHSETSLDLK--VDKEKKELEVTVVIFSMNE 1403
QY 599 -----LEDSNDGREDNDISREFKSDILNDVSGTSDI----- 631
DB 1404 VGTSDAQAEFGHEHTPCSSSEIKDESOGSEE---SVEVSKETVOGESSEKDVNMLDVO 1460
QY 632 --IGKYGNSSEITTKTAPPSDNNK-----ENSKSLDPANNESL-----OO 675
DB 1461 SGSEKYYQENEPDISLVS---KTENGDKFEELIPSVEGAGLDETHNQLDLDESIVKO 1516
QY 676 OLEVPTKEDDSILANNSNIAPPELTLPVNEANDYSSNDVTKTDAVSPEE----- 729
DB 1517 SLDTPEEETSKTIDKIEDKKEEVTLLH--QOEREGSGYGLTKD--EAVSVLESRELGEQ 1574
QY 730 -----SLSREHETDSKPINFTISIMHOKOKKHQIHKVPTKOILASUYOQKNEQ----- 778
DB 1575 POOEELCLANQENETK-----LQEOVDKHE---PTKEEVSNDGQSPVEELSNEY 1622
QY 779 -----BSRVTSKV---KIPNAIOFKKFKEVNWSRRVVSPPMDOLNVSQF 821
DB 1623 IOVSSASLSEGREYETVVAEAKIGEOVADKIQ--KSFEGELIVEAHSSLPSSSEKEHET 1681
QY 822 LPELSEDSGFKULNFANYSNNTNRRPSFTPLSTKNV--LSNIONDPVNEPREPKSAELR 880
DB 1682 VSEKTDDEKVAE---PIGDMREGLDIAETHTLSLSVQDKEVDDEIHPSVALPID 1737
QY 881 NARRLSANKAAPPLPPLPQOPSPSTRSNKRVSRFVPFETIRTSALAPCDMYND 940
DB 1738 EQEKTSTEGFTKSGSEARDKRDDEHVDSSTPMLEKNDNENQTSKTSEDV--CMQOE 1795
QY 941 IPDDGAGSGPTTKAGMKTLPSSMDKDVKRLNAKKGT-----ODEVINAKIV--- 990
DB 1796 -----SGTLEVKPBEEKSKDSOEISETIEIEATSDQTLPIETSHNTLISSELVSEQ 1849

QY 991 -DOKPKKNSIVTDPEDRYEELQOTASTIHNTIDSSIVGRPDISITDMLPYLSDELAKP-- 1047
DB 1850 DDQSPKKEVEIHHEEPKEADVATSERNLVPETS--DADNTLSSQLVSETKREHKLQAG 1907
QY 1048 ---PT-----ALLS-----ADRLFMQOEVLNPLRSNVLVHPGA 1077
DB 1908 EILPTLEIPRESSDEALVSLASREDKVALQDNCADVDRETNDIOEBSISVETEESV 1967
QY 1078 G-----AATNSMLPEPDEFELINSPARVNSNSDVAISGASTISFN 1120
DB 1968 GETKPREHEDEIDAVHTPTARILLENSESLINEAK-----KONEELNETEKYVALD 2022
QY 1121 QLDNFDQDA-----TIQKIOEQPASKSANTVRGDD---GLASAPRP-----PT 1164
DB 2023 HEEEPVNHAPKLEETKDEKSOEIPETAKATETITDQTLPIGTSQADQTPSVLSDKDDQT 2082
QY 1165 PTKKESISSKPAKLSSAPKSPFIKIGSPRVYIKKNGSIAGTPIKATHK----- 1215
DB 2083 PKOVEI-----LEETKETHKVOAEDIFST 2108
QY 1216 ---PKKSFQNEISN--HKVRDGISPSG-----SEHQHNPMSVSPSOYTDTATSTVPD 1266
DB 2109 ETVPKESFTIAPVSMLASGDEPVTPOEGDYANTQOEHR---VSAET----- 2153
QY 1267 ENKDVOHKPREKOKKHNNHHNNHKKOTDIPGVYDE-----IPDVGLO----- 1312
DB 2154 EKVVGETKRESQAE-----GAEKSD--DQVEDESKTKTVDVAGLENDYPTBEA 2201
QY 1313 ERGKLFPRVYIGIKINILPDIINTHKGRFTLLDNGVCHVCTTPREVM-----DDH 1360
DB 2202 EHDEYVSTLPVVGIT--LTQLOT-----TLETALINDASSEVSMKLEPADEKKGDV 2255
QY 1361 NVAIGKEFELTVADSLFLLTLKASYEKPRGTLVETEK----- 1399
DB 2256 VESNEKOF---VSD-----ILEAKRLHGDKSGAEKIKESGLAGKSLPTEINLOEENK 2308
QY 1400 --KVVASRRNLSLRFSKDIITTKFVPT-----VKDTWANKFADGSE---ARCTI 1447
DB 2309 EVKVOETREIAQVLRREELISSLSPLSAEQENHVISDEKOEEREPQOOFNGSTSEKISL 2368
QY 1448 DLOQFEDQITGKASQFDLQCFNEMETMSNGO-----PYMKRGPKYKA----- 1490
DB 2369 QVEMHLKOFETSKKEQD---ETHETVKEEDQIVDIKKDKKDDDEDEIVSSEVKKKNKA 2424
QY 1491 -OLEVKNLVPRSDPRELLPTSIKRSAYESINELNBNQNNYFEGYTLQOEGDGPFFKKRPF 1549
DB 2425 RELEVGNDRVSRDGEKEVRPHNALENEEEMNEVASEK-----QISDPYGVYKKA-- 2474
QY 1550 KLMGTSLLAHSELSHKTRAKINLSKVVDLIVYDKENIDR-----SHHRNFSOVLLLDHA 1604
DB 2475 -----SEAEHE-----DPVDDI---KSNDDDFPTQOAPKDDSDSEVSADETV 2513
QY 1605 -----KIKFANGELIDFCAPRNK-----EMWIQNLQOEITYRRNRFRQ 1643
DB 2514 PKRAIGEELKAVSSKVLDDIOENSNTAEVATNFADRLPQONLSLQISQNSRQ 2567
RESULT 8
Q26216
ID Q26216 PRELIMINARY; PRT: 2771 AA.
AC Q26216:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE Rholety protein.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN (1)
RP SEQUENCE OF 379-2771 FROM N.A.
RC STRAIN=YM;
RX MEDLINE=97077455; PubMed=8920022;
Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;

Db 1386 ITIQLKDLLKQOQOQOENNKE-----IERLQIEQLKQOQOQOENLSELSNKEIK 1448
Qy 1329 LPDITHKGRFTLTLDNGVCHVTPEYNNMDHNAIGKEFELTVADSEFILTAKASYEK 1388
Db 1449 IQ-----TTQOEFQOLSHN---RSKDQHLQOQOQOQOQOQOQOQOQOQOQOQO 1486
Qy 1389 PRGTLVEVTEKKVKSNNRL-----SRFGSK-DITTTKFEVPEYKDTMANKRPAED-- 1439
Db 1487 Q-----DHQFKKVIDERYNLQOLEQSTLSNNQDLQLEKRLPLELDSNEKQKTIDLL 1541
Qy 1440 GSFACVYDLOQOFEQITGKKSQF-----DLNCFNEMVTMNGNQPMKRGKPYKIA 1490
Db 1542 SNISMLQISLQNDKDLSEIRNSNITLESRTQQLSLDEKDLKDLQOQOQOQOQOQOQO 1596
Qy 1491 QLEVMALVPRSDPREI-----LPTSIRSAVESINELNNEONNPFEGYLHOEGDCPIFEK 1546
Db 1597 QOPFASSPSSPSSSLSTPTPKQRPNOIEIDLVLVEIVNRNDLIRKN-----KT 1649
Qy 1547 RFEKLMGTSLSLAHSEISHKTRAKINLSKVVDLIYVDKEN 1585
Db 1650 KFYKLENGDIYVNSIIT---YRLSLDDNDSDLIQOYEEN 1685

RESULT 11

Q9NFS3 PRELIMINARY; PRT; 16215 AA.

AC Q9NFS3;
DT 01-OCR-2000 (Tremblrel. 15, Created)
DT 01-OCR-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE D-Titin.
GN SLS OR D-TITIN OR CG1915.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Y.Q., Broadie K.S.;
RT "Characterization of Drosophila D-Titin gene."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AJ271740; CAB93524.1; .
DR HSSP; P56776; ITLK.
DR FlyBase; FBgn0003432; sls.
DR InterPro; IPR002106; AATRNA_ligaseII.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001452; Ig_MHC.
DR Pfam; PF00047; Ig_50.
DR Pfam; PF00018; SH3_1.
DR SMART; SM00408; IGC2; 15.
DR SMART; SM00410; IG_Like; 34.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
DR PROSITE; PS00002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match 3.9%; Score 337.5; DB 5; Length 16215;
Best Local Similarity 17.0%; Pred. No. 2.1e-05;
Matches 286; Conservative 287; Mismatches 656; Indels 453; Gaps 71;

Qy 41 TNSKPSLDPNSSDPTTSEQDEKKEKKTAFOTSFDRNFDLNSIDI-----QQ 92
Db 12173 TTYVETPRDDQPSVKRRTKRIKKIKKDEVDVFKRVIDEEAQPQPGSVLVIEDVFPKP 12232
Qy 93 TIOHQO 145
Db 93 TIOHQO 145

Db 12233 SSEKRRKKPKIKDKNHTSVEEETPRHEDEVLIESVPEDSPLSDLLITVVDVSPRIEPEENKV 12292
Qy 146 PVIYDMSPKSITMKATPRASPCKVAFVY-----INPEIHNPDRNV 187
Db 12293 NOIETD---KKPEKKKKRPSAKTILEENVEDTVEKPLEALHTSDLEKPDQOESISTK 12349
Qy 188 EEEEDSOOKEDSVPEPLIOWKMDPSQFNYSDEDTNASVPPTPLHTTKPRFAQKLNKN 247
Db 12350 EEEQKHTPEKKKSSKISEQKQKSTEGYELSVTEHDKP---EEKPRFYVQIQSET 12405
Qy 248 EV-----NSPEEA-LTDMKLRKRNFS 270
Db 12406 NVEETKDDTGVKVKQVTKRMLRRPAGEGELEIEVVRDDQPEAEITVEVEPE---PVN 12462
Qy 271 IDEKYNLVLSPTNNNSKNVSDMSHLOWASKNKTN-EIHNLSFALKAPKNDIENP 329
Db 12463 ODEKK---EPKKTTRKAKKDDIHDTYQKLELTPKTELEVEKIEFEPIYKDKPLSP 12519
Qy 330 INSLTNADISLRSSGSSSSL---QSLRND---RVLESVP-----GS 366
Db 12520 IDVLDESPEVQKKKSKSTICKGOCVEEAPQEPQILEVAPVEVDVKEVITEDGK 12579
Qy 367 P-----KVNPGSL-----NDGKGSDEVVESLPRDLSRDKLETTK 405
Db 12580 PVQKTTKRVLKIKIGPERQTTFKITMIESDNDSTVAVDEPEIASQSI-----E 12631
Qy 406 EHDAPENHNENFIDAKSTNTNKGQLVSDHLDSEFDSYNHTQESILNLKASASOIS 465
Db 12632 EH--PEQSEKELAPKPKTVKR---VKKD-----LSDYVKKLEIEEP 12670
Qy 466 LNALEKOROTOEQTOAEPBEETSFSDNIRVKQEPSNLEFVKVTKKEPVSA-TEIK 524
Db 12671 KYDLKEKYEVE-----MPEKPVKLTVSDSIPEEPRPD-----KSGIISVLPDTT 12714
Qy 525 AKRFESSRIILIKMED-----ELAEPAD-----IHKKKNE 556
Db 12715 KPKKTKTPKTPKTEDTDOQVDEPETTYVDTDIDELPTQTAQEDATQAOTPSAOE 12774
Qy 557 ANSHVEDTDALKALNDEESDTQNSTKMSIRPHIDSMKLEDSNG--DREDDDIS 614
Db 12775 KSTQ-DQTKDITQKIVKHKKTPTDQKSVESISELPEVHKDQVQIITHELVEEQPEKIT 12833
Qy 615 REFESDILNDVQSQSDT---IDKYGNSSEITTTTLAPRNSDNDKENSLSLEDPANNE 671
Db 12834 EVRVIDEAVEEESQPIVEEVEDEPOPATETVEDVTKPS----- 12875
Qy 672 SLQOOLEVPHTEKEDDSILANSNINAPBELTLPVVEANDYSFNDVTKT-PDAYSFPES 730
Db 12876 ---KKKKVKKKTDH-----DELKKMLE-----QELKTELEKYEKIEFD 12914
Qy 731 LSREHETDSKPINFTSIWHKQOKKHQIHKVPYTKQIIASYOQYKNEQESRVTSKVRIP 790
Db 12915 VPKKTLPEFAALEPIKIERKEQ-----PTKVITL-----DATQVP 12950
Qy 791 NAIQFKKK-----EVNV-----MSRRVYSPMDLNSQFLPELSEDSGFOLDNFA 837
Db 12951 KTVKLPKSRKREKPAEELTVOLPKFRKLARMV---LVEPPAPLIPKTTDIDGAIKD--NG 13005
Qy 838 NTSNNTNPR---SPTPLSTKNVLSINIDPNVPEPEPKSAEIRNARNSANKAANQ 894
Db 13006 ELSRNIAGEAEELKPKPKTKTI-KIKDDLEKVLKYEKY-----ISSEEPPEEK 13056
Qy 895 APPLPQROPSSSTRNSNKRVSFRVPTFEIRRTSSALAPCDMYNIDFDFGAGSKPTIK 954
Db 13057 TPYKKPEKAPK-----PEEKQDQVLLKLGKGGKKKKE 13088
Qy 955 AEG-----MKTLPMDKDDVKRILNAKKGVTODEYINAKLYDQKKKSIVTDPDRYEE 1009
Db 13089 EEAPEVNLKINIPKQPO--EVEEEVELKQKPEVEIVEEQ--TKKPKDEFFVEP----- 13139
Qy 1010 LQQTASINNATIDSSITGRPDSITDM---LPTYSDELKRP-PTALLSADLMEQEVH 1064
Db 13140 -----FEPSEFDPEIVPELQOIEHPLEIKVKKPSKTKYKPKDKSKSEPE-- 13186

```

QY 1065 PLRNSVYLVHAGCATNSMLPEPDEFELINSPARNVSNNSDNVAISGNASTISENOL-- 1122
D 13187 -----TIVSELVAGVPRKEEALPEBODVK-FRKPREDAPREED-----SEIKLRPRPQ 13222
QY 1123 ---DMNFDDQATIGOKIOEPASASANTVRGDDDLASAPETPRTPRKESISSPKARLS 1179
D 13233 ASKENDPEQALVTPRK-AEPIPIQIEEDKALIDDEKPKRKPQKQPEQIEIAKEEPEEF 13291
QY 1180 SASPRK-----SPKIGSP---VRVYKKNNGSIAGIEPIPKATHKKKSPQGNELIS-NHKV 1230
D 13292 EVASKEEBALVDKPLEIEKPRVDKVEKKRPEAPVSEVVVIEEERKPREVEPEIEVEYKI 13351
QY 1231 RDGSGPSSEHOOHNSMVSVSQYDATSTVDEKNKDVONKREKOKONHNHNHNH 1290
D 13352 TTYVLEPERDAP--KEHOKYKVIDFERQETTEEVI--EEKVVTTRKKRPPQRODEEEVNLK 13407
QY 1291 HHKOKTDPGVVDEIDPVDVGLQERGLKFRVGLIKINIMLPDINTHKGRFTTLONGVHCY 1350
D 13408 EPKEEQIQPDVVSAL--SLPIEE-----PEQKPRQYEVLELKITG----- 13445
QY 1351 TTPRYNMDDHVAIGKEPELVAVDSL--EFLITLKASYEKP--KRTLYEVTEKKVYKSRN 1406
D 13446 TTPPE-EPNDVQIAVKEKVKTKPVKKVKEDKLVVEAEKEKQPEETIVEV-EKQEKKRS 13503
QY 1407 RLSSLFGSKDITTTTKFV---PTEYKDTWANKFAPDGSFARCYIDLOQFDDQITGK--AS 1461
D 13504 EKPKSYERK--ISFTQSTIEEKPIEVAEE--APEET-----PKVEKVKVAKRPSY 13549
QY 1462 QFDLNCFNEMETMSNGNOPMK-----RSPKYKIAQLEVKMMLVPPSPREILPTSIIR 1513
D 13550 EFTLKEPDEEKVITVDDQPEEAPVVEVFKKKRPEAVEAEFV--MTEPIVETSVSE 13606
QY 1514 SA 1515
D 13607 TA 13608

RESULT 12
Q91704 PRELIMINARY: PRT: 6815 AA.
ID 091704:
AC 091704:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG18242 protein.
GN CG18242.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Goezanne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butlis J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry K.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasner K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```

```

RA Harris N.L., Harvey D., Hejman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Melnikov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paolel J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -i- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AE003473; AAG22226.1; -.
DR HSSP: P56276; 1TLK.
DR FlyBase: FBgn0035301; CG18242.
DR InterPro: IPR003962; F0111.repeat.
DR InterPro: IPR003961; FN.IIT.
DR InterPro: IPR003598; Iq_C2.
DR InterPro: IPR003600; Iq_1like.
DR InterPro: IPR003006; Iq_1MC.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; iq; 11.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00014; FNTYPEIIT.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGc2; 5.
DR SMART: SM00410; IG_1like; 6.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
KW Immunoglobulin domain; Repeat; SH3 domain.
SQ SEQUENCE 6815 AA; 779559 MW; A4E244001A4EBA01 CRC64;

Query Match 3.9%; Score 336; DB 5; Length 6815;
Best local similarity 16.9%; Pred. No. 8.5e-06;
Matches 286; Conservative 283; Mismatches 660; Indels 466; Gaps 69;

QY 41 TNSKPSLDPNSSSDYTTSEODEGKEEKKDPAFQTSFDRNFDLNSIDI-----QQ 92
D 2740 TTTVTEPRPDQDQPVKQKRTKKIKKDEVEDVKVRIEEARQPRESSVDLVYEDFVRKP 2799
QY 93 TIOHQOQDPOQOQOQLOSDNDN-----LIDFSFQTPMTSLDLTKONPYDKNENHIA 145
D 2800 SSEKRRKKRPIKDKHTSVSEETPHEDEVLIESVPEDSPSLDDLITVVDVSPPIEEENRV 2859
QY 146 PLYINTSPKSIIMKATPRASPCKVAFV-----TWPEITHYPNRV 187
D 2860 NOIETD-----KREKKKKRPPKAKILIEENVPEOTVEKPLFALTDSDLKRPDQERSISIK 2916
QY 188 EEDDSQCKEDSVEPPLIIOHQMKDPSQFNYSDEDTNNAVPPPTPLHTTKPTFAOLINKN 247
D 2917 EEOQHHTPEKKKSSKISSEQKQPTSEQYELISVTHDLK-----EEKPFTVOYIQSPT 2972
QY 248 EY-----NSEEDA-LTDMKLKRNFNSNIS 270
D 2973 NVEETKDDTGKVKHKOVTTRKMLRRPAGEGEIEIIEVVRDQDEPAETIIVYEPE--PVN 3029
QY 271 IDEKTNVLVPTNNNNNSKVNVSMDSHLONLOASKNKTN-ENIHNLSPFLAKPKNDIENP 329
D 3030 QDEKPK---EPKKTKRKKAKDDIHDIYIOKLIELETPKTELEKYEKIEFPPIYKXPLDSP 3086
QY 330 LNSLTNADISLSSGSSOSLOSLRND-----NRELVSVGSP----- 367

```

Db 3087 IDVLDESPKVEYOKKDKKSRSTKVPNEETPVQEOYAKVNVVEEAPDEPILVOILEYKPV 3146
Qy 368 -----KVNPGSL-----NDGKSGEVESESL 392
Db 3147 EVDYKEVITEDEGKVPQEKTTKRVLLKIGPEQOTFKITIMIESEDNDSYTVYVDEPELAS 3206
Qy 393 PRDLSDKLETTKHDAPENHNNENFIDAKSTNTKGLLVSSDHLDSFDRSYHTEQSI 452
Db 3207 POST-----EEH--PEOSKEKLAAPKKTIVK--VKDD-----L 3237
Qy 453 LNLNSASQOISLMALEKOKOTQOEQTOAEPDEETSESDNKKVQEPKSNLEFVKVT 512
Db 3238 SDVYKALLEETIPKVDLEKEVE-----MPEKPVKLTVSDISPEEPKPD----- 3282
Qy 513 IKKEPVSA-TEIKAPKREFSSRLIKNE-----ETAEPAD 548
Db 3283 -KSQPIVLPPTTKPKTKPTKPTEDTDQOVPEDETTVDTPIDELPTQAOED 3341
Qy 549 -----IPKKENANSIVEDTALLKALANDESDTTQNSTKMSIRPHIDSKLEDSDN 603
Db 3342 TATAQITPSAOEEKSTQ-DTKDTIOKTIVKHKTKPTQOKSVETSELPVEVAKDYQISIIH 3400
Qy 604 DG--DREDNDISREFKSDILNDVQSTSDI--IGDKYGNSSSETTTKTLAPRSDNDX 658
Db 3401 EELVEEOPKEILEVRVIDEVAEVEESQPIVEEVEDPEQATEETVEDYKPKS----- 3455
Qy 659 ENSKSLDEDPANNESLQOOLEVPHTKEDDSILANSSNIAPEELTVPEVANDYSFNDVT 718
Db 3456 -----KKKKVKKKTTD-----DELKKMLE-----GEIE 3481
Qy 719 KT-FDAYSFEESISREHETSKRINFTSIHKOEKOKHIIHVPKTOIIASVOQYKNE 777
Db 3482 KTELEKTEKIEEDVPKKAKEPFALEPIKIERKOK-----PKVTTL----- 3524
Qy 778 QESHVTSQVYKIPALIOFKFK-----EVNV-----MSRRVSPMDLANSQFLPE 824
Db 3525 -----DATVPRKTVKAKPKRKEKRAPEELTVQLPKRLARV--LVYEPAPRLIPK 3574
Qy 825 LSEDSGEKDLNFANYSNNTNRP--SEPLSTKVNLSIDNDPVNEPPEPKSYAETRN 881
Db 3575 TTDTGALIKD--NGELSRNIEAEELIKFKPKTKTI--KKIKDDLEKVELEKYEK----- 3626
Qy 882 ARRLSANKAANOPARLPOROPSTRSNSNKRVSFRFVFPFELIRTSALAPCOMYNDI 941
Db 3627 ---ISSEPEPEKTPYKKEAPK-----PEKQEDV 3655
Qy 942 FDDGAGSKPTIKAG-----MKTLPMDKDVKRLNAKKGVTODEYINAKLVQKPKK 996
Db 3656 KLLGKGKKKKKEEAPENVTLKNIPQKQ--EVEEVELKQKPEVELVEEQ--TKRKPD 3712
Qy 997 NSIYTPEDRKEELQOTASINNATIDSSITGRPDSISTDM---LPLYSDCLKRP-PTAL 1051
Db 3713 GEFVPER-----FEPESEFDRPEVVPDELEOIEHPELPEKVKPKSKTKY 3755
Qy 1052 LSAORLFMEQEVNHLRSNLSVLNPGAGAAATSSMLPEPDEFILNSPANNVSNNSVNAIS 1111
Db 3756 KPKKKSSEPR-----TIVSEIVAGVAKKEEALPEQDV-FRKPREDADBEDT----- 3802
Qy 1112 GNASTISFNOL-----DMNFDDOATIGQIOEPASKASANTVRCDDDLASAPETPRPT 1166
Db 3803 ---SEIKLRYRQASKDNPREQALVTPK-AEERIPQIEKALIDDEKPKKSKPKKQOP 3858
Qy 1167 KKESSISKRAKLSSASPRK-----SPIKIGSP--VRVYKNGSIAIEPIPKATHKPKK 1218
Db 3859 KEOGLKEBEREFYSVEEALVDKPIEIKPKDVKKKPKKAPKPEAPVSEVVVVEEPRK 3918
Qy 1219 SFQGNES-NNKVADCGISPSGSEHQOHNSWVSQYTDATSYVDEKKNQOHKPRE 1277
Db 3919 EELVEEELPEVEKITTVELEPAD--KEHOVKVIDFEROETTEVI--EKKVUTRKKKP 3974
Qy 1278 KOKOKNNNNNNNNNNKOTDTPGVVDDIETPVQLOERGKLFRRVLGININMLPDTNTHKG 1337
Db 3975 KPOOREVEVTLLKPKKEQIOPDVVSAET-SLPTEE-----PEOKPEQY 4017

Qy 1338 RFTLLDNGVHCVTTPREXNMDDHNAIGKEFELTVADSL--EFTLLTKASYEK--RGTL 1393
Db 4018 EVELKINQ-----TTP-EPRNDVOLAKVEKTKPVKVKEDKIVVVEAEERKQVETI 4071
Qy 1394 VEYTEKVKYKSRNRLSRLEFSKDIITTTKVV--PTEYKDTWANKFAPDGSFARCYIDLO 1450-
Db 4072 VEV-EKOEKSKSEKPKSYERK--ISFQSTIEEKPIEVAEE--APEET-----PK 4116
Qy 1451 QFEOITGK--ASOFDNCFMEWETMSNGNOPK-----RGKPKYIAQLEVKMLYVP 1500
Db 4117 VVEKVAEKFPDSTYETFLKEIDEKXVITYDDQPEEAPVEVFKKKPKPEVAEVEV--- 4173
Qy 1501 RSDPREILPTSIKSA 1515
Db 4174 MTEPKIVETSVETA 4188

RESULT 13
Q9NDS4
ID Q9NDS4 PRELIMINARY; PRT: 2678 AA.
AC Q9NDS4;
DT 01-OCT-2000 (Tremblrel, 15, Created)
DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
DE AMIB.
GN AMIB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=20118106; PubMed=10651904;
RA "Kon T., Adachi H., Sutoh K.;
RT "Amib, a novel gene regulated for the growth/differentiation transition
RT in Dictyostelium".
RL Genes Cells 5:43-55(2000).
DR EMBL; AB030033; BAB01489.1;
DR InterPro; IPR00194; ATPase_a/bcentre.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
SQ SEQUENCE 2678 AA: 298921 MW: 7082BFBFBEBETC7CA CRC64;

Query Match 3.98; Score 335; DB 5; Length 2678;
Best Local Similarity 18.18; Pred. No. 3e-06;
Matches 288; Conservative 228; Mismatches 605; Indels 472; Gaps 65;

Qy 34 KPLNPRNTSKPSLDPNSSDPTTSEODQEGKEKKTATQTSFDRFDDNSIDIQT 93
Db 142 KPINNSSNNS-----NITSSITSDSLKRLKKSPESTTPPNTNNN---SNVTKDSP 192
Qy 94 IOHQOQPOQOQOOLSQT-DNNLIDEFSEFQPTMTSLDTLTKQNPV-----DKVENHAPT 147
Db 193 PNATKMSSTPSKLSPTISNNNNNTTAAATTTTNTNNSNSPSTNNNNNNNNNSPS 252
Qy 148 ---YINTSPKNSIMKATPKASPKKVAFTYTNPRELHNHPDKRVEEDOSQCKEDSEVE-- 202
Db 253 SHNVVNSPP-----STSSRSPPVAVSTSN-----TSTIOPIS 285
Qy 203 PLIQHQKDPQSFVNSDSDTNAVSPPTPELHTKPTFQQLNKNNEVNSEPEALDKMLK 262
Db 286 PLINRQ--TSSHNTQO-----POQPYVNHOP-----LQPIQIETTYTELW 323
Qy 263 RENFSNLSDEKVNLYLSPNTNNNSKNVSDMD-SHLQMLDASKNKKTENIHN--LSFA 318
Db 324 VFNISNTSL-----LDFSHLNVEGIVNLDNIGNSVYIYOYL 362
Qy 319 LKAPKNDIENLNSLTNADIS-----LRSSGSSQSSLOSLRND--NRVLESVPGSPKKV- 370
Db 363 FKSLYNLIVE--YNNVSNDFIKIGNSFKSHILOKOKS-RNDYNNSTISGVSCKRLF 419
Qy 371 -----NPGLSLNGICIGFSDEVEVESLLPRLSLDKLETTKEH 407

SQ SEQUENCE 3111 AA: 349635 MW: EDA9A5FD38115773 CRC64;
 Query Match 3.9%; Score 333.5; DB 5; Length 3111;
 Best Local Similarity 17.7%; Pred. No. 4.2e-06;
 Matches 274; Conservative 252; Mismatches 595; Indels 427; Gaps 65;
 27 STFNSTKLANPRPNRSKSLDNPSSSDTYTSEDOEKKEKKOTAFOTSFDRNFDLN 86
 991 SFNNSLTQKMSSESSSKTPID-----LKDDEPRSSSSSK-----N 1026
 87 SLDIOOTIO-HOQOOPQ-----OQOLSOTDNNLIDFSPOTPMSTSLDITKONPTVD 138
 1021 QKDSEKTLKHSSEPSSTKETTGETEASDNDKIGETKEPQIKKLPSTAE--LED 1084
 139 KYVENHAPTYI-NTSPNKSIMKATPKASPKYAFVTNPETIHVPDNVEEDSOQKE 197
 1085 RFNALERKMSVQKSSPSKN--KKEPDEESK--STKEPE-----PPESEKANEKT 1132
 198 DSEVPRLLIOHKKDSQFNYSDEDTNNAVPT---PPLHTTKPTPAQLLNKNVEYN--- 250
 1133 SGRQPTPIAKKSDSDQ---KKSETKENQSPTKNDEKVKVSPKSEEMIEKETSSNPKE 1189
 251 ---SEPEALTDKIK--RENENSLDEKYNLYLSPNNNSKNVSDMSHLONLQDASKKK 307
 1190 DSHESAATNKKEVEGRRELSSEKGDHKIKEKSEAPGKAGKETAEYKN--ANYKDSKKG 1247
 308 TNEINHNLSFALKAPKNDIENLNLTNADLSRSGSSOSSLOSLRDNRYLESVPGSP 367
 1248 DSQ-----KNEAKTSVSQTESD--LKPSKENSSTKQAEQCK-----TP 1285
 368 KAVNGLSLNDIGKGSDEVE---SLPRDLSRDKLETTKEHDAPENHNEPDAKSTN 424
 1286 RNSP-----STEELEKRFNALKOMSTNLLETKEPDDQK-----PAIK 1325
 425 TNGQGLVSSDDLSPFDRSYNHTEOSIINTLNSASQSOISINALEKOROTQ---QEO 481
 1326 SOSTSEVYTKQMSKSFDKIKEVNAV---EKQSRVEVNAEKKKKNVEAPRANKG 1382
 482 QAEPREETSFSDNIKVKQEPKS--NLEFVYTKIKKPYSAATEIKAPREFSSRLIRKN 539
 1383 DSQEPESQHKGNQRRASEPSTEDLEKRYETLKRMSKNOFSTVDEALERT---Q 1438
 540 EDEIAPRADIHKKEKNEANSHEVDLALKALNDESDTTONSK---MSIRPTISD 596
 1439 QVYISEAVN---EKKRPSTEDLESREALHGDKNVESKMDKEKHVDAVIEAIPIS- 1492
 597 WKLESNDGREDNDISFEKSDILNDVQSOTDIIGDKY--GNSSEITTKTLAPRSON 655
 1493 ---PPPPPPPKERYLA---EPVLHQOALIELOSKMRQSGEENLK---PSEIN 1541
 656 NDKENSKSLEDP-----ANNESLQOOLEVPHTKEDDSILANSSNTA 696
 1542 PQRRQKLLQRPMPGDETSEAPANTAYYRAANHEQOQRM-----VRRSDLP 1590
 697 PPEEL--TLRVYANDYSSFNVDYTKTFDAYSSFEESLSNEHETDKPTIFISIMIKQEO 754
 1591 SNADELNRIQLEFOLYKFF-----YKORCADCSEVAAR----- 1624
 755 KKHQIHKVPYTKQIIASYOQYKNEOSRVTSDVKIPNAIOFKKFEVNVMSRVVSPD-- 812
 1625 ---VKLPREDPSTSRQAKQBAEQLEQRYLALEKQISEMSKLKLEMRERHSADDS 1680
 813 -----MDDLNVSOFLPEISEDSGCFDLNFA-----NYSNNTNRPRST 850
 1681 GSPRLSTETIDATGKELVRYTQNGILEEVDANHPINISIKIMVNDKSSQKQKGS 1740
 851 PLSTKNVLSINDNPVYVPERPKSAETRNARLSANKAAANQAP-----LPPOQ 903
 1741 KPTEDLRLRLE-----QLEQOLLEERAKNGSIPPENEVLEEKREKLE 1783
 904 PASTSNNKRVSRPRVPTFEIRRTSALAPCDMYNDIIDDGAGSKPIKAEGKTKLPS 963
 1784 EKDSCKKQKKNCHNQHVKGDEVEKTE---IPADKRIE-----PASKET-----KTL 1829

964 MDKDYK-RILNAKGV-----TODEYI-----NAKLYDOKPKNSIYTPED 1005
 1830 VEAQGRRAVVDTEKSVKQGNVATDEKSVQDDNNVYDQKADKRLDKK-----SPAG 1884
 1006 RIEELOQTASINHAATDSSITGRPDSISTDMLPYLSDELKCPPTALLSADRLFMEQVHP 1065
 1885 KSEDTKQTS-----CKKEK-----SEDIKQASEA----- 1908
 1066 LRSNSVLVHPGAATNNSMLPEPDELINSPARNVSNNSDVVAISGNASTISPNOLDN 1125
 1909 -----PKAGASKETRGKPRSTKLEKPTYESVUKETPRPKENLES----- 1950
 1126 PDDQATIGQKIOEPASKSANTVRGDDGLASAPERTPTPKESISSPKALSS--ASP 1183
 1951 -----EKPKSKNEATK-----TEQKSKERT--VAVSPRESKVSQKOME 1990
 1184 RKSPIKGS-----PYAVIKNGSIAIGETIPKAT-----HKPK----- 1218
 1991 KKETIKDSSSKELPEKMYINSTDYGMPPNGKTVVLLMNEHRASKVVRRLTRANTELED 2050
 1219 SFOG-----NEISNHNKVRDQ---GISPSSGSEHOQNPMSVSPQYTPATSTVDEN-- 1268
 2051 LFOALEKQNDKRLVKSSEGLRIVDPKPSAQVEOTQAIISDTKETEIDFTSAKPEENP 2110
 1269 ---KDVQHKPREKOKQKNNHNNHNNKQKT-DIPGVYDDEIPVGOERKLPFRVYG 1323
 2111 KEAKEDKREPEREDFDMGRPTVKHNLKRTVYLPSTKELSRFSRLQIKL----- 2164
 1324 IKINLPDINTIKGRFTLLDNGVHCVTTPYNNMDHNAVATGEFELYVADSLFPLTLK 1383
 2165 LEDVERKIDVEORL-----NEIERKIKLOYSL-SHEKOLKLYLELCBGKGLD----- 2209
 1384 ASYEKRGTLVEYTEKKVYSRNRRLSRLEGSKOIITTKFVTEYVUDT 1431
 2210 --DDEVPRVETPTKEAETITADR-SRSPGRKALATKSPYTSPPSKAT 2254
 RESULT 15
 Q17464 PRELIMINARY; PRT: 3147 AA.
 AC Q17464: Q22160; (1996-01-01, Created)
 DT 01-NOV-1996 (EMBL) 01, Last sequence update)
 DT 01-MAR-2002 (EMBL) 20, Last annotation update)
 DE T04F3.1 protein.
 GN T04F3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
 OC Rhabdilitidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases;
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 Craton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sproul J., Wohldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.

Ra Kershaw J.;
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 274026; CA98419.1; JOINED.
 DR EMBL: 272513; CA98419.1; JOINED.
 DR EMBL: 272513; CA98419.1; JOINED.
 DR EMBL: 274026; CA98419.1; JOINED.
 DR InterPro: IPR004839; AminoTransf1/2.
 DR Pfam: PF00155; aminoTransf1/2.
 DR SEQUENCE 3147 AA; 357854 MW; A97D05E4FE6A79 CRC64;
 Query Match 3.8%, Score 331.5, DB 5, Length 3147;
 Best Local Similarity 18.8%, Pred. No. 5.2e-06;
 Matches 349; Conservative 283; Mismatches 666; Indels 561; Gaps 82;
 QY 2 NSTPS--KLIPDKSHLOLOPOSSASIFNPTKPLNPRKNSPSIDPN----- 50
 DB 422 STLPFRHTAPVD-----LELE-----DIFN--PKPFSHPANSKPPPPNRRRHPPSAS 469
 QY 51 -----SSDVTY 57
 DB 470 VAHSSSFVDESPQHFVTTTIDRNQVTPVTTTTNNMRSGFLNHSNOLSSDNNV 529
 QY 58 SE-----ODQKKEKKDPAFQTSFD--RNFDLNSIDIOQTIOHQOQOPOOQOOLSGT 110
 DB 530 AQMMMDASEMENNVSSRKSILSTSSTSARAKASVARISVDLTGPKRKSQAVALPSS 589
 QY 111 DNNLIDERSFOTPM-----TSTLDLTKON--PTVDKYN--ENHAPTYINTSPKNSIMK 160
 DB 590 DVSIDIDTALTLPKREYQSEKRTFIVTKORADVDALDFEKDRTP----- 638
 QY 161 ATPKSPKVAFTVNPINHYPRDNVREEDOSQOKEDSEVEPLIOHQKDESOQFNYS-- 218
 DB 639 APPRGKRRKSDATENOPETILELVEKHEIDSSKYSTSTIN--LNDEMEFRNTNDS 695
 QY 219 --DEDFTN-----ASVPTPLHTTKPTF--AQLKNNVSEPALTDMLKRP- 264
 DB 696 SFDEVNPRNOLVYEIPFSEPRVTSTATIOLESSDVAGENSE--KRP--VISMSSKSEI 752
 QY 265 -----NFSNLIDEXVNLILSPTNNNSKNVSDMSHQNIO 301
 DB 753 AKKEDAOAGCVIIPHSKHEILDESINISMDVFN-----TPHOCRAVPIDA----- 801
 QY 302 DASKKNTN-----NIHMLSFALKAPKNDIENPLNSLTNADISL----- 340
 DB 802 DSKKTSDDRSVSTVITINDNV--PTEEPK-----LVAKNCELEAEERIKRIKOF 854
 QY 341 --RSSSSSSSLOSLANDNRVLES-----VPSGPKVNPGLSLDNGIKGF 383
 DB 855 ERTTGEQEISKSSEPTEDEMPDEKDHRTSAVSLDKVFVQGTAKKPE-----ND---EF 906
 QY 384 SDEVVESILPRDLSDKLETTEKHA--PEHNNENFIDAKST-----NTNKGOLLVSSD 436
 DB 907 DEKIRGIAEFERSKQKEVORSVAEYSHSGKHLPDESINISMDVFNFSQ--KYKSD 963
 QY 437 HLDSEFDSRYN--HTEOSILNLSASQSOQISLALAKOROTOEOQTOAEPEETSFSN 495
 DB 964 KLSPTERTVEPEVSTATNMLDNILIASGATIR--BEENTVLEBEERIKQVREEFKTTEN 1021
 QY 496 IKVQOEPSNLEFV--KVTKKEPVASATEIKARREFSSRLIKNEDE-----IAEPADI 549
 DB 1022 LEIOKEVVLTKEEVNSDVKEHRTSAVNIDLEKVFTHGSSKPKPKNDIRGIAFEERT 1081
 QY 550 HKKKENANSHEVDALLAKKALNDEEDDTQONSTKMSIRPHIDDMKLEDSNDDRED 609
 DB 1082 KOEKAQRSTVLE-----TSOSNS-----RIFEESISMDVFNLSLN 1120
 QY 610 NDISRFEKSDILNDVOTSIIIGDKYNSSEITTKTLAPRSDNNKENSLSL----- 664
 DB 1121 ESQVSEITFASPDPDLVLTSTFHNVIIEKIDDDVTKT-----DSNVEEKKQVALRID 1174
 QY 665 --EDPANNESSLOOQLEVPHTKEDDSLANSNIAPEELTLVPEANDYSSFNDVTKTFD 722
 DB 1175 EFKRPTTEONLOKEFLTKKEEYSV-----KMEKRTSAVSIDLKVPD 1218

QY 723 AVSFEESLSREHETDSKPIINFISIMHKOEKOKKHQIHKVPTKQIIASVQYKKNBESRV 782
 DB 1219 -OSSKETTYS--NETDEKIKRGIAEFERSKQKEVORSVAE---TSHSGKHIFDESNI 1271
 QY 783 TSDKVKIPAIQOFKKEEVNVSRRVYSPDMDLANSOFLPELSEDSGKDLNFAVYKNN 842
 DB 1272 SNDDVF--NTSQKXKDEKLSSEPTVEPEVSTATMN--LDNIIFASG-----IATREEN 1322
 QY 843 TN-----RPSFTPLSTKNVLSNIDNDNVVEPEPPSYAEIRNARLSANKAAP 892
 DB 1323 TVLEEEERIKQVREEF-----KTTENLEIOKEVVLTKEEVNSDVKEHRTSAVNID 1377
 QY 893 NQAPPLPQROPSPRNSNKRVRVPTFELRRS-----SALACDMN----- 939
 DB 1378 DVF-----IQSSKHPEDEDEKIRGIAEFERSKQKEVORSVAEIVTSQSNKHIFDK 1431
 QY 940 -----DIPDFGAGSKPTIKAGMKT--LPSMDKDVK--RILMAKGYTODEY--INA 987
 DB 1432 SNISMEVFNESQNGKDSNIDMKETDMPKEKRDQRYVDVHROKPPENGFEPTFNG 1491
 QY 988 KLYDOKPKKNSIVT-----DP-----ED-----RYELOQTASTIHMT 1020
 DB 1492 SKISNEPKQISITTLNDNVFTEEPKLVAEDECEIEAEERIKRIKOFERTTGOEIL 1551
 QY 1021 IDSIIYGRDSDISTDMLPYLSDELKPRPTALLS--ADRLFMEQEVNPLNSVLYHPAG 1078
 DB 1552 KNS---EPAEDT-----SDE--KKNHTAAVSIDLKVFVO----- 1582
 QY 1079 AATNSMLPEPPELINSFARVNSN-----NSDVAISGNASTISFNOLDKNPD 1128
 DB 1583 -----GTAKKRPND--EPDEKIKRGIAEFERSKQKEVORSVAEYSHSGKHIFDESINISMD 1638
 QY 1129 QATIGOKIOEPASANTVRCDDGLASAPETPRPTTKESISSKPAKLSAPKSPSP 1188
 DB 1639 VENTSQKVK-----SDEKLSPTERTVPR-----EYSTATNMLDNI 1673
 QY 1189 KIGSPYRVIKNGSINGIEP--IPKATHKPKKSPOGNEISNHKV--RDGGSPPSGSEHO 1245
 DB 1674 IFASG:ATEKNTDVEEERIKQVREEKTTENLEIOKEVVLTKEEG---DNSDVKD 1729
 QY 1246 HPSNVSV--PSQYTDATSTVDENKDVQHKRP-----EKOONHNRHNNHKKOCTDI 1298
 DB 1730 HKASAVNIDLDFVIGRSSKHPEDEDEKIRGIAEFERTKOE-----KEARSTY 1781
 QY 1299 PGVVDDELIDVGLQERKGLFPRVLGIKINLPIRINHKGRFTLLDNGVCHVTPREYNMD 1358
 DB 1782 -----IEQYSSKMF-----NESDISDVVFNTSQKSDSEKLSSEPTVE 1823
 QY 1359 -----DHNVAIGKEPELTVADSL-----FILTAKSYEKP--GLVETYEKK 1400
 DB 1824 PEVSTATNLDNVALSKEKKNENNETOEEDIOQVREEFKSTEOKIOKIELTKEE 1883
 QY 1401 VVKSRRNLRLFGSKDIITTTKFPVTEVKTANKPAPDGSFARCYIDLQOQEDQITGKA 1460
 DB 1884 CSDEKELKTYSGS-----IDDKVFIQSSSK 1911
 QY 1461 SQFDLNCFNEMETMSGNOMPKRGKYKIAOLEVYKMLVYPR---SDPRILPTSRSAV 1516
 DB 1912 PRND---ESDERINGIAEFERTKOEKEAORSVUVVETSPSNKHISDESSISMDIEFSRS 1967
 QY 1517 ESINELNNEONNFFEGYLQOEGDCCPFKKRFRKMGTSILANSEISHKRAKINISKY 1575
 DB 1968 Q-----DNKSTSNFE-----KSGSIFP-----IVLDGEKEVASASINLNGV 2004

Search completed: March 17, 2003, 12:27:14
 Job time : 206.719 secs

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:20:59 ; Search time 32.8137 Seconds
(without alignments)
4875.027 Million cell updates/sec

Title: US-09-964-858-1
Perfect score: 8631
Sequence: 1 MNSTPSKLLPIDKHSILQ.....WVNLMLQ0000000SSSQ 1664

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8631	100.0	1664	T18216	integrin-like prot
2	696.5	8.1	1320	S57113	BUN4 protein - yea
3	367	4.3	704	T50303	hypothetical prote
4	346.5	4.0	2401	T28676	rhopty protein -
5	342	4.0	2364	A56577	microtubule-associ
6	338.5	3.9	5327	T13564	microtubule-associ
7	337.5	3.9	1738	T14867	interaplin - slime
8	331.5	3.8	3147	T18674	hypothetical prote
9	329	3.8	6713	B89921	hypothetical prote
10	325	3.8	1650	T18444	hypothetical prote
11	324	3.8	1302	JC6009	surface-located me
12	320.5	3.7	1381	S45781	probable calcium-b
13	314.5	3.6	3724	T18427	hypothetical prote
14	314	3.6	2464	1 ORMSPI	microtubule-associ
15	313.5	3.6	3488	T34418	hypothetical prote
16	313	3.6	1658	S55101	hypothetical prote
17	312.5	3.6	1819	T32008	hypothetical prote
18	310	3.6	1875	S38173	myosin-like protei
19	309.5	3.6	1271	A45555	glutamate rich pro
20	304	3.5	2346	T13829	Tpt homology - fru
21	299.5	3.5	2269	T28677	rhopty protein -
22	297	3.4	2271	T90073	hypothetical prote
23	296	3.4	2139	T18296	myosin heavy chain
24	293.5	3.4	2481	D90011	FmbB protein [limo
25	293	3.4	5170	T15348	hypothetical prote
26	292.5	3.4	1790	S67593	transpirt protein
27	291	3.4	1744	JH0720	transin - African
28	290.5	3.4	2663	S28261	centromere protein
29	290	3.4	3343	T42207	breast cancer susc

30	285.5	3.3	1979	2	C71622	hypothetical prote
31	285.5	3.3	3869	2	A48205	All-1 protein +GPE
32	284.5	3.3	1727	2	T50073	myosin-like coiled
33	284.5	3.3	1937	2	T38077	hypothetical coile
34	284.5	3.3	4688	2	F82885	hypothetical prote
35	283	3.3	1364	2	T40839	hypothetical prote
36	282	3.3	3924	2	S37431	ankyrin 2, neuona
37	281.5	3.3	2722	2	T20532	hypothetical prote
38	281	3.3	1358	2	A28360	Sir4 protein - yea
39	280.5	3.2	1233	2	S56271	hypothetical prote
40	280.5	3.2	1435	1	BVBYL1	guanine nucleotide
41	280	3.2	1139	1	E64234	cyathadene-acces
42	280	3.2	1466	2	A36426	SPA2 protein - yea
43	277.5	3.2	1189	2	S56852	hypothetical prote
44	277.5	3.2	1621	2	A82255	hypothetical prote
45	277.5	3.2	3225	2	I52300	giantlin - human

ALIGNMENTS

RESULT 1									
T18216									
Integrin-like protein alpha chain - yeast (Candida albicans)									
C:Species: Candida albicans									
C:Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 11-May-2000									
C:Accession: T18216									
R:Gale, C.; Finkel, D.; Tao, N.; Meinke, M.; McClellan, M.; Olson, J.; Kendrick, K.;									
Proc. Natl. Acad. Sci. U.S.A. 93, 357-361, 1996									
A:Title: Cloning and expression of a gene encoding an integrin-like protein in Candida									
A:Reference number: 206510; MUID:96133936; PMID:8552638									
A:Accession: T18216									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-1664 <GAL>									
A:Cross-references: EMBL:U35070; NID:q1144530; PID:q1144531; PIDN:AAA96019.1									
A:Genetics:									
A:Gene: alpha INT1									
Query Match									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MNSTPSKLLPIDKHSILQPOSSASIFNSPTKPLNFPRTNSKPSLDNSSDITYSQ	60						
DB	1	MNSTPSKLLPIDKHSILQPOSSASIFNSPTKPLNFPRTNSKPSLDNSSDITYSQ	60						
QY	61	DOEKGEKKKDTAFQTSFDRNFDLNSIDIQOTIQHQOQPOQOQLSOTDNNLIDFSG	120						
DB	61	DOEKGEKKKDTAFQTSFDRNFDLNSIDIQOTIQHQOQPOQOQLSOTDNNLIDFSG	120						
QY	121	QTPMSTLDLTKONPVYKVNENHAPTYINTSPNKSIMKATPKAPKVAFTVTPETI	180						
DB	121	QTPMSTLDLTKONPVYKVNENHAPTYINTSPNKSIMKATPKAPKVAFTVTPETI	180						
QY	181	HPDNVVEEDSOOKEDSVEPPLIOHOKKDSQFVNSBEDTNAVPPPLHTTPETA	240						
DB	181	HPDNVVEEDSOOKEDSVEPPLIOHOKKDSQFVNSBEDTNAVPPPLHTTPETA	240						
QY	241	QLLNKNEVNSEPEALTDKLLKRENFSLDEKVMYLSPNTNNNSKNVSDMSHLQNL	300						
DB	241	QLLNKNEVNSEPEALTDKLLKRENFSLDEKVMYLSPNTNNNSKNVSDMSHLQNL	300						
QY	301	QDASKKKTENIHNLSPALKAPKNDIENPLNSTNADISLRSSGSSQSLQSLRNDRYL	360						
DB	301	QDASKKKTENIHNLSPALKAPKNDIENPLNSTNADISLRSSGSSQSLQSLRNDRYL	360						
QY	361	ESVPSGPKVNVGLSLNDIKGSDVEVSLPRDISRKLETTKHDAPENHNEFTDA	420						
DB	361	ESVPSGPKVNVGLSLNDIKGSDVEVSLPRDISRKLETTKHDAPENHNEFTDA	420						
QY	421	KSTNTKGLLVSDHLSDFRSYVHTEQSILNLNSASQSLSNALAEKOROTEOBO	480						
DB	421	KSTNTKGLLVSDHLSDFRSYVHTEQSILNLNSASQSLSNALAEKOROTEOBO	480						

Db	421	KSTJMNKGQILLVSSDHDJLDSFSDSYNHTGSDSLIANTLNSASQOSISJMLALEKROQOEDQ	480
Qy	481	TQAAPEBEETFSUNIKYVKQDPRKSNLEFYKVTTIKKEPVSAITEIKAPKEFSRRILIKNE	540
Db	481	TQAAPEBEETFSUNIKYVKQDPRKSNLEFYKVTTIKKEPVSAITEIKAPKEFSRRILIKNE	540
Qy	541	DEIAEPADIPHKKEENANSVVEDTDALTKKALNDDEESTQONSTKMSIRFIIDSDMKLE	600
Db	541	DEIAEPADIPHKKEENANSVVEDTDALTKKALNDDEESTQONSTKMSIRFIIDSDMKLE	600
Qy	601	DSNDGREDNDISIREFKSDIILNDVSQTSDIIGDKTGNSSSEITTKTAPAPSDNNKEN	660
Db	601	DSNDGREDNDISIREFKSDIILNDVSQTSDIIGDKTGNSSSEITTKTAPAPSDNNKEN	660
Qy	661	SKSLIEDPANNEISLOOQOLEVPHTEKEDDSILANSSNINAPPELTLPLVEANDYSFNDJKT	720
Db	661	SKSLIEDPANNEISLOOQOLEVPHTEKEDDSILANSSNINAPPELTLPLVEANDYSFNDJKT	720
Qy	721	FDAYSSEESLSREHETDSKPINFISIMHKOEKQKHQIHKVPTKOIILASYOQYKNEES	780
Db	721	FDAYSSEESLSREHETDSKPINFISIMHKOEKQKHQIHKVPTKOIILASYOQYKNEES	780
Qy	781	RVTSDKXKIPRALQOFKKFKREVNMSRRVVSPPMDDLNLVSOQLPELSSEDSGFDLNFAYS	840
Db	781	RVTSDKXKIPRALQOFKKFKREVNMSRRVVSPPMDDLNLVSOQLPELSSEDSGFDLNFAYS	840
Qy	841	NNTNRPRSEFTPLASTKNVLSIINDPNVVEPERPEKSYAEIRNRNRILSANKAANQAPRLP	900
Db	841	NNTNRPRSEFTPLASTKNVLSIINDPNVVEPERPEKSYAEIRNRNRILSANKAANQAPRLP	900
Qy	901	QROSSSTRMSNKRKVSFRVPTFEIRTSALSALPCMYUDIFDDBGAGSKPTIKAEKMT	960
Db	901	QROSSSTRMSNKRKVSFRVPTFEIRTSALSALPCMYUDIFDDBGAGSKPTIKAEKMT	960
Qy	961	LPSPDKDQVRLILNAKKGVODEYINAKILVDQPKKNSIYTPREBYREELQOOTAISHNT	1020
Db	961	LPSPDKDQVRLILNAKKGVODEYINAKILVDQPKKNSIYTPREBYREELQOOTAISHNT	1020
Qy	1021	IDDSITGRPDSISTDMLRYLSDELAKRPYALLSADRLFMEOEUNIRLSNSVLNBRGAGAA	1080
Db	1021	IDDSITGRPDSISTDMLRYLSDELAKRPYALLSADRLFMEOEUNIRLSNSVLNBRGAGAA	1080
Qy	1081	TNSSMLERPEBELINSAPRVNNSNSNVAVISGNASTISFNOLDIMNPDQATIGQITOPOR	1140
Db	1081	TNSSMLERPEBELINSAPRVNNSNSNVAVISGNASTISFNOLDIMNPDQATIGQITOPOR	1140
Qy	1141	ASKSANTVRGDDDBGLASARETPRTPTKKESISSKPAKISSASPRKSPITIGSPVAVIKNN	1200
Db	1141	ASKSANTVRGDDDBGLASARETPRTPTKKESISSKPAKISSASPRKSPITIGSPVAVIKNN	1200
Qy	1201	GSIINGIERIPKATINHKPKKSTQGMELISNKKVYRQDGISPPSSGSEHOONHSMVSPOUTDA	1260
Db	1201	GSIINGIERIPKATINHKPKKSTQGMELISNKKVYRQDGISPPSSGSEHOONHSMVSPOUTDA	1260
Qy	1261	TSTVRDENKQVONKRYEKOJKONNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	1320
Db	1261	TSTVRDENKQVONKRYEKOJKONNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	1320
Qy	1321	VLGIKATINLRDINTNGKRGYLLDLNDGVCSTUTREVMMDHNAVIGKEBELVYADSLEFLT	1380
Db	1321	VLGIKATINLRDINTNGKRGYLLDLNDGVCSTUTREVMMDHNAVIGKEBELVYADSLEFLT	1380
Qy	1381	TLKASYEKRGTLVEVUEKVVUYSRYNLSYLSFESKQIITTTFRVREYVOTMANKEAFDG	1440
Db	1381	TLKASYEKRGTLVEVUEKVVUYSRYNLSYLSFESKQIITTTFRVREYVOTMANKEAFDG	1440
Qy	1441	SFASCYUIDLOOFEDQITGKASQDFELNCFMEMETMSNGNPMKRGKRYKIAQOLEVMKLVPR	1500
Db	1441	SFASCYUIDLOOFEDQITGKASQDFELNCFMEMETMSNGNPMKRGKRYKIAQOLEVMKLVPR	1500
Qy	1501	RSDBRELPLTINSAYESINELNNEQONNFVEGTLNDEGSDCPTKKRPFKLMGTSLAHS	1560
Db	1501	RSDBRELPLTINSAYESINELNNEQONNFVEGTLNDEGSDCPTKKRPFKLMGTSLAHS	1560

```

OY 1561 EISHTTRAKINLSKYVDLYVDKENIDSRNHNFSVLLLDHAFKIKFANGELIDCAAN 1620
Db 1561 EISHTTRAKINLSKYVDLYVDKENIDSRNHNFSVLLLDHAFKIKFANGELIDCAAN 1620
OY 1621 KHEMKIWIIONLOEIIYRNFRFRQPVWNLMLQOOOOOOOOSSOO 1664
Db 1621 KHEMKIWIIONLOEIIYRNFRFRQPVWNLMLQOOOOOOOOSSOO 1664

RESULT 2
S57113
BUD4 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein J1905; protein YJR092w
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 29-Oct-1999
C:Accession: S57113; S61952
R:Ramezani Rad, M.; Kirchhath, L.; Hollenberg, C.P.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57111
A:Accession: S57113
A:Molecule type: DNA
A:Residues: 1-1320 <RMB>
A:Cross-references: EMBL:Z49592; MIPS:YJR092w
R:Sanders, S.L.; Herskowitz, I.
submitted to the EMBL Data Library, November 1995
A:Reference number: S61952
A:Accession: S61952
A:Molecule type: DNA
A:Residues: 'MIDAESTVSLKEIDNEMEQKSNITONGSEDTPIHMKLPLQEIQDQTMELVKHINTSRNATENSGRGSRP
202-212,'D',214-1007,'A',1009-1320 <SAB>
A:Cross-references: EMBL:U041641; NID:g1151116; PIDN:AA017116.1; PID:g1151117
C:Genetics:
A:Gene: SGD:BUD4
A:Cross-references: MIPS:YJR092w; SGD:S0003852
A:Map position: 10R

Query Match 8.1%; Score 636.5; DB 2; Length 1320;
Best Local Similarity 20.9%; Pred. No. 1.5e-21;
Matches 346; Conservative 262; Mismatches 567; Indels 483; Gaps 69;

OY 137 VKV-NENHAPYITMSPKSIKMKATPKASQKVAFTVNTPEIHHPNREE----- 189
Db 5 IDKLARDEKRPVKLSSPLKFLTKSTQPLS-----YPSPIHRSIETE 49
OY 190 ---EDSOQKEDS---VEPRLIOHWKMDQSFNYSDSDTINASVPRPRLHTTKPTFAQL 242
Db 50 TNYDEDEDEEDDAVYCLTQSPQILHS-----PSRIPI-----TNAVS 86
OY 243 LMKNEVNSPAPALDMKLKRENFNSLSIDEKVNLYLSTNNNKSQVSDMSHLQNLQD 302
Db 87 INK-----LNLDFTLN-----PNESDKSLVSD-----TSVDS 113
OY 303 ASKNKTNENIHLNLSFALKA-PK-----NDIENPLNLTGADISLRSSGSSQSL 353
Db 114 TGRELDFTIPLPLPCMSSTREMTPVDEKCNLPKSLTLNNSHSDSRANSGLCGGFKNF 173
OY 354 RNDNVLVESVQSPKPKVNBGLSNDGKIGFSDEYVBSLLRPLSRDKLETTKEHDAREIN 413
Db 174 NNESSVLL-----IPAKIIQSLMLR-----MRLKLTMLCRDLQNM-----EHI 211
OY 414 NENFLIDAKSTN---TNKGGLVSSDDHLDSPDRSVNHTQSLTLNLSASQSLNMLE 470
Db 212 DEAFPEKKVLDGCSNRPVTLGEND-----TRSIYVS-----NKGTANV----- 252
OY 471 KQRTQEQEQTOAAPEEETSFSDNIKYQKPKSNLEFVKVITIKKEPVASATEIKAPKEF 530
Db 253 --QFSQEDSLASHSPKFKDLNATSDDVWNDEKETDANISTSKSPESYIADYKWTROE- 309
OY 531 SSRILIRINNEDIAPRADHPRKKEANEANSHVEDTALLKALNDDEESDTQNSTKMSIR 590
Db 310 -----DWDTRKLHQESHAHQEPRIIPQKQSSSEPTTELNNESEFORN 352

```

```

QY 591 FHIDDMKL---EDSNDGDREDND-----ISREKSDILNDVQSOTS---DIIG 633
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 353 FKGDEYRIVQHEESLYGQRTKSPENIINSSEIGDHEAPENPLAKTAGEHDLS 412
QY 634 DKYGSSEITTKTLAPPSDNNKSKSLDPANNESLQOOLEVPHTKEDDSILANS 693
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 413 SCEODSVSEARKKDRIEKEVEETKDENETEKDESEYHKEVEENEDEHP----- 462
QY 694 NIAPP-----EELTL---PVVEANDYSFN-DVTKTPAVSFPESLSREHETDQPINFS 746
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 463 -LLPPLPMEETQFNEPFIIDENDTSDSLITRSM-----KPSDIYS 503
QY 747 IMHKOQKQKHQIHVPTKQIIASVQOYKNEQESRYTSKV- KIPNAIQKFKKVVNM 804
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 504 IMHIDEEETKSN---SPESIASNQ- FSOQSSITTAIVSCKKNGSTSEFKRRIYS 556
QY 805 SRRVVSPPMDILNVSOFLELSEDSGFKDLNANFNNNTNRPSTPLSTKVNLSIND 864
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 557 RSRIYNPKSRVSLVNYDNE-----DYILNSEMNALDPMRRNTLISKRIODNIQTQ 608
QY 865 -----PNV-----VEPEPKSYAEIRNARRLSANKAPNQPAPLPQR 902
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 609 KCHAPLIRPSIMKLNGEDSGFONHLEVPQOEHENIPLSTHLEODITTVNG--LDEQK 666
QY 903 QPSSTRSNKRVSRFRVPTFEIRRTSSALAPCDMYNDIFDFGAGSKPTIKAECKTLP 962
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 667 LPTNTODEAE-----ISIREESA-----GDI--TFNRDLDLSL----- 698
QY 963 SMDKDDVKRILNAKKGVTODEYIMAKLVQDKPKNSIYTPDEPDREYELQOTASIHATID 1022
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 699 SFDEELGQFANFLDALDHD-----SISFNHGPDD 728
QY 1023 SSIYGARPOS-ISTDMPLYLSDCLKKPT-----ALLSADRLEMEQDVNHLRSN 1069
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 729 SSSFOSSKSKFNSLWESSYELKPPSTIRKQPIADVLQKLESTK--DDADLEKIRRE 787
QY 1070 SVLVHPGAGAAATNSMLPEPDELINSAPRANVSNNSDNVAISGNAS--TISFNQDL---- 1123
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 788 RI-TERPRGLGIG-----MLKTPYKDV-----IALASIKGYEASPSDTSRPE 831
QY 1124 -MNPDDQATIGQIOEOPARSKSANTYVRGDDGLASAPETPRPTTKESISSKPAKLSSAS 1182
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 832 GANNSDATILNM-----FDDFEEDKMTPTPVR--SI----- 861
QY 1183 PKKSPK--IGSPVRYIKKNGSIAGIERIPKATHKPKKSFQGEISNHNKVR--DQGISP 1237
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 862 ---SPIKRVSSPFEVVK-----AGKKQENNEINIKAEERIEP 896
QY 1238 SSGSEH---QOHNPMSVSPQYTD--TSTVRDENKDVQNHKPREKOKNHNHNNHN 1291
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 897 MTOQETDGLKODIRPLLAOTKDNVEAKETITQLEERD----- 935
QY 1292 HKOKTDIPGVNDELIPDVGLOERKLFVRVLGIKINILPDIINTHKGRFTLLDNGVCHYT 1351
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 936 -----VQEFEPDM-----GTLYLISIKAIISTLALYGTSHRATYVAIVFNGENVQ 980
QY 1352 PREYVMD-DHNAVIGKEPELV-----ADSL-EPILLLKASYEKREPTIVE 1395
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 981 TWESLPTDGNIRINKEFELPIDKGTETSSASSERDSYKCVTLKCKYERPRHLEV 1040
QY 1396 VTEK-KVYSRRRLSRFLFESKDIITTTKFEVTEVK-DTWANKFAPDSEFARCYID--QOF 1452
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1041 IYDKYVPVGS-----FFCKTKYKFEKKYVQKKPKDDEMDYLFACQSGFARCEIETNEER 1094
QY 1453 EDOITGKASQFDLNCFNEMETMS--NGNO---PMKRGKPYKIAQL-EVKMLYVRSDDPR 1506
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1095 LKNVAFNFSHMHYNMINKMSRIADKIHGSKRLYTELRPKARPKHVAASLDVACFELERTSAFE 1154
QY 1507 ILPTSIKRSYESIINELNNQNNYFEGYLHOEGGDCP-IRKKREPKLMGSLASHSTISK 1565
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1155 QRPKQPSLVNKKVSKKLOONITKEGYLLDGGDLGKGIENRFFKLHGSLGSGYHEISRK 1214
QY 1566 TRAKINLSKVVDLIYDKENIDRSNHRNPSDVLLLDHAFKIKFANGELIDFCAPRNHEMK 1625

```

```

Db 1215 AKIDINLKVTKVTLRNMEDIQADONGGGRNFMDWLFHBCFQLVDPDEBRITFNMECSNEEK 1274
QY 1626 I-WIONLQELIITRNRRRROPVWMLLQOQOQOQOQSS 1662
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1275 SDWYNKLOEVELNVE-HQPMWAKYCEKLAIEEKTRPTT 1311

RESULT 3
T50303
hypothetical protein SPAPYU7.03c [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T50303
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Saunders, D.; Harris, D.
submitted to the EMBL Data Library, January 2000
A:Reference number: Z25059
A:Accession: T50303
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-704 <MCD>
A:Cross-references: EMBL:AL136521; PIDN:CAB6312.1; GSPDB:GN00066; SPDB:SPAPYU7.03c
A:Experimental source: strain 972h(-); clone plasmid pYU7
C:Genetics:
A:Gene: SPDB:SPAPYU7.03c
A:Map position: 1

Query Match
Best local similarity 21.0%; Score 367; DB 2; Length 704;
Matches 189; Conservative 140; Mismatches 315; Indels 256; Gaps 38;

QY 787 VKIPNAIQKFKK-----EVNMSRRVVSPPMDILNVSOFLELSEDSGFKLNFANYS 840
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 21 LRIPLPTDTDTCCDYASTIASISREKTRMRNRSNISTAFSAESEDADGDFPD 80
QY 841 NNTNRRSF-----TPLSTK-----NVLNIDNDPNVPEPEPKSYAEIRNARRLSA 887
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 81 QLTSSNSFDDHQSLLPSTREVRRTPTYSVMNTESSSTVE-----DVAKENILSL 132
QY 888 N-----KAAPNQPPLPPOKOPSSSTRNSNKKRVSRFRVPTFEIRRTSSALAPCDMYNDIF 942
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 133 NDSCLIKLSDEA-----SNKS--SRSTPRNISTKSNSS----- 164
QY 943 DDFGAGSKPTIKAECKTLPMDKDDVKRILNAKKGVTODEYIMAKLVQDKPKNSIYTD 1002
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 165 -NOGHGDIPL-----PKN----- 177
QY 1003 PEDRYELOQTASIHATIDSSIYGRPDSISTDMLPYLSDCLK-KPTALLSADRLEMEQ 1061
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 178 -----PARVCNSKL-----FNEDTLPAEFEEVISIPVKL-----EL 210
QY 1062 EYVHLRNSVYLHPGAGAAATNSMLPEPDELINSAPRANVSNNSDNVAISGNASTISFNQ 1121
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 211 PTH-----SHNSS-----DTSFTNS--IYSSVSDWVGLGEGINSIA--- 244
QY 1122 LDMNPDQATIGQIOEOPARSKSANTYVRGDDGLASAPETPRPT--KKEISSKPAKLSS 1180
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 245 -SRGFSEDSFODITKTPPLSFADENR-----ECKRDIYRSDSIHIEERPLIS 293
QY 1181 ASPKPSPIKISPVRYIKKNGSIAGIERIPKATHKPKKSFQGEISNHNKVRIGSISPSG 1240
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 294 -----SITSLDSP-HVLDEN--APIPLPKVVSPL----- 320
QY 1241 SEHOQNPMSVSPQYTDATST-VPDENKDVQNHKPREKOKNHNHNNHNHKKOKTDIP 1299
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 321 -----DPRTNLSAFDALTRTYLLRONSKVVAHNASOKQEMOTSRVNVNSCWYPES-LS 373
QY 1300 GYVDDDELIPDVGLOERKLFVRVLGIKINILPDIINTHKGRFTLLDNGVCHVTTPREYVMD 1359
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 374 KNISSSLQGTG--GSGRLVRLMEIRNLITPLASGMITRTTYIT--SGKH-IQVP-WNALH 428
QY 1360 HNAVIGKEPELVADSLFELITLKASYEKPRGLTVEYTEKKVYKSRNRLSRFLGSKDIIIT 1419
| : : : : | : : : : | : : : : | : : : : | : : : : |

```

Db 429 STTKLENEYTFEESISSIVCTLRAYADPK-----VRRSTLGKVFST---N 473
Qy 1420 TTKFVPE-VKDTWANKFAPDSFARCYIDLOQFEDQITGKASOPDLNCFNEM--ETMSN 1476
Db 474 KRKSMITDVSALHGFSESDGTFGEBVTINTDSVSRALGRCQSVLPIMKMTYDPAK 533
Qy 1477 GNOPKRKPKYIAOLEVKMLYPRSDPRET--LPTSTRSATESINELNEONNFFEGYL 1534
Db 534 DVKPL----PRKVGLELHVFLP-ALPYSLEKLPASIESAMYLKLAEMDRTLDCDYL 588
Qy 1535 HOEGDCPIFKRPFKLMSTSLASHSEIKTRAKINLSKVVDL----YVDKENIDRSN 1590
Db 589 CQGGDCCPYWRRYFOLLGSKLVAFOQPSKVRATIDLEAHYHYDDNHSDEELE--- 645
Qy 1591 HNFSDVLLDHAFFIKFANGELLDFCAPNKHMKIWIQLOEIIYRNRFRQPVNML 1650
Db 646 ----GYLFESGFRIFSGNDYIDFYAETVGEKDEMMSTLQHIGQCSMVHKMTKSL 700
RESULT 4
128676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676; A45521
R:Simha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: 220507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:91041784; PID:91041785; PIDN:AB41263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cc
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281
Query Match 4.0%; Score 346.5; DB 2; Length 2401;
Best Local Similarity 18.9%; Pred. No. 1.3e-06;
Matches 360; Conservative 327; Mismatches 736; Indels 477; Gaps 94;
Qy 12 DKHSLDLOPOSSASIFNSPTKPLNFPRTNSKPSLDPSSSDYTSBEDQKKEE--K 69
Db 603 DKTELETRFETGLSNHESNNKELLYFYDLKANLGRKKNMLYKQFNEKEKAVEIDIK 662
Qy 70 KD-----TAFQTSFDRNFDLNDSIDIOQTIOHQOQOPOOQOLOSOTDNMLIDE 117
Db 663 KKVVDINKIYSNLEITITYSIYNEDTEN--EIGKSIELNTRYLEKAYANTNLNEIKE 720
Qy 118 ---ESFQT-----PMTSTL--DLTKQNPYVDK-----NENHAPTY--- 148
Db 721 KLKDYDFDFGFEKKNIKYFDEKNKIDNIDTLQKIDKSIETLTLEIKNSENHIDEIKGI 780
Qy 149 --INSPKSIKKATPRASPKKAVTY---TNPEIHYPNRRVEEEOOQKEDSVAP 202
Db 781 DKLRKVPNKTME--DPEKEIKIENIYEIKDKKNIKYKEIDLKLEISKISINDTSLF- 838
Qy 203 PLIOHOMKDPDSOFNS-----DEDTNASVPTPLHTTKPTFA-----QLL 243
Db 839 -----KLKININISYKSLGNLPLQOIDEKKKAE-----HTIKMEATIIDLDNIK 884
Qy 244 KKNNEVNSPEALTDMLK---KRENFSLSLDEKYNLYLSPTNNSKNVSDM--DSHLON 299
Db 885 KKSQIEKEMINMDIKMDIHKEMKALNISHD-YKIY-HTTSKHEEKISDIRNSLKI 942
Qy 300 LODASKNKTENIHNLSFALKAPKNDIENPLNLSLTNADISLRSSGSSGSSLOSLRDNRY 359

Db 943 IODFSEESYINDIKR-----ELEKNVLESQNN---NTDIN-----QYLSKJININYI 986
Qy 360 LBSVSPKPKVPLGLNDNGIKGFSDEVES--LLPRDLSRCK--LETTKEHDADEH--- 412
Db 987 L-----KLKIKIKIIDKVEYETDEIEKNKKKINAELSNEKEITITQKESSLKECOS 1038
Qy 413 ---NNENFIDAKSTN-TNKGQLLVSSDDHLSDFDRSYNHTQGI-LNLN---SASQS 462
Db 1039 KIKSTIDNYSVECIKNTITNLKTYIYNEKNININTFYKNAEEYNQVSLNFNNIEADTKS 1098
Qy 463 QISLVALEQRTQOQEOBT--QAAPEETS-FSDNI-KVQOEPSNLEFV-----KTI 513
Db 1099 QYTLN-IKKNGTNTNDVNIKELKHKKSNYKQDAGKNTQEIKNKLEFKEYQEVTV 1157
Qy 514 KKEPVATEIK---APKREFSRILR-IKN-EDELAEPADIRPKENE-ANGSHVEDTAL 567
Db 1158 LNKYVALELKKKPKDKTKNYSQIITKEIKDAHNTTTSQADKSEKKNELKNOIIEBEV 1217
Qy 568 LKKALNDEESDTQNSTKMSIRF-----HDSQKLEDSNDGDR-----DNDDISR 615
Db 1218 AKNNKSNKALIDIQLSVEPFKIKFLKIDLPRTKSDCKLETQDIEFKISLSDTQETKL 1277
Qy 616 FPKSDIWDVSGTSDIGDKYNSSEITTKTLAPRSDNNKENS--KSLDPANNESL 673
Db 1278 IENKNILNTELEKLSLKNQKKNIEDQ--KREL-----DEVSKIKITIESNVQHK 1327
Qy 674 QOOLVPHTKEDDSLANSNSTAPPELTLPLV-----EANDYSFNDYTKTFDAYS 726
Db 1328 NWEIGIVE-KIMEIKAKANKDOLESQOKLIPRIKILNISFKANDLEGI-DTKNKLGXNT 1385
Qy 727 FEESLSREHEDSKPINFISITWHKQOKOHOIHPKTOIATYQYKKNEDSVNTSDK 786
Db 1386 ---ENKNNTYEIEIKSYDILT-----HYLETVSKEPI--TYEQIKNK---RITQON 1427
Qy 787 VKIPNAIDQKKPK-----EVNVMGRVY--SPMDQDLNVSGFELPESDSGFKOLNAN 838
Db 1428 ELTITAKVNRKKSFLDIDEANEPRIYTHFNKKLADVA-DKFTNEYSK---VKKGDN 1482
Qy 839 YSNNTNRPSFTPLSTKYNLSNIDNDPVNVEPPEKSYAEIRNARRLSANKAPNQAAPL 898
Db 1483 IENSNINNVKST---DENLLNLIHQTK-----EMVANIYSKKYYSKYKAEINIFINI 1532
Qy 899 PRQOPSSSTRSNKRVSRFVYFTEIRTSALAPCDMYNDIFDFGAGSFTTKAEGM 958
Db 1533 P-----KLANSINLQI-----KSSSGI-----DLFKNINIALIPLYDSQK 1568
Qy 959 KTL---PSMDKDDVAKRLNAKKGVYODEYTNAKLVQDKPKKNSIYTDPEDRYEELQOPAS 1015
Db 1569 DTLTPIPSPEK-----TSEYI--TKISDSIVTLLDILKRSOELQKQOALN 1613
Qy 1016 I--HNATIDSSIYGRPDSISTDMLPYLSDELKPPALLSADRLFMEQ--EYHPLRSNS 1071
Db 1614 LLEFNRLHLDKYQ-----ATNELKDITLSDLKKNKKEOILNKKYLLHKSNEILKSCNS- 1666
Qy 1072 LVHPGAGATNSSMLPEPDFELINSAPRVNSNSDNVAISGNASTISFNQDMFDDQAT 1131
Db 1667 -----QNYDTLIESSKYDKIKESNNYEKEKENLGI-----NEDVKA- 1703
Qy 1132 IGOKIOEQPASKASANTVRDODGLASAPETPTPTPKKESISSPKALS-----SASPRK 1185
Db 1704 ---MEQFNNDIKIIEKENNYKHSEKDNYNFSENNNILOSKKRLKLELNAFNAEIK 1759
Qy 1186 SPIKIGSPRYVIRKKNGSINGI-----EPKATNHPKK 1218
Db 1760 IEDKI-----IEKNGLINKLLETKKOCMLFTYKTLVETLIKITDYTKFTISAKRFSKE 1813
Qy 1219 SFQNEISNKKVRDGISFSGSEHQOHNPSWVSPQYT-----DATSTVPD- 1266
Db 1814 FLKYIDATSNLNDINTLOTKYDLNQLINKHVAADATNDNNMLIEKREKATITINML 1873
Qy 1267 -----ENKQVO-----HKPRKQOKOHHRHHH-----HHKQKT 1296

Db 1874 TELFTIDSNKIDADGLHNKIQIIFYNSELSHSIKOLYKKHMAFKLLIGHINKKYF 1933
Qy 1297 DIRGVVD-----DEIPDVG--LOERGLKFRVIGIKININLPDINTHNG---- 1337
Db 1934 DISKFEFDNLOESELATNLNDLKEIGOKISDKKQFLLAH--SETPIPNNTLKETIYH 1991
Qy 1338 ---REFTLNDGVCHVTTPPEYNNMDHNVAGKEFEFTVDSLEFILTILKASYEKPRGTL- 1393
Db 1992 DIVAKKNQIDE-IEMIT-----NEENENITLYMDIITKLKMKKQSIINFPYTTENDSNIIT 2046
Qy 1394 -----VEVTEKKVYKSNRLSRL---FGSKDITITTTTKFVPEYKDTW 1432
Db 2047 QHIQNNENDVYSKIKESLETTIQSFOKLINKEIKAQFYDNNNNINISTISQVDVY- 2105
Qy 1433 ANKFPADGSAFCYIDLOQFEQITGKASQFDLNCFNEMETMSNGNQPKRGPYK--IA 1490
Db 2106 KKHISKDLTENELIQIKSLEDI--KKSTYDI-----RSBQIRKYVNPPIHDVYE 2153
Qy 1491 QLEVMLYVRSRDPREILPTISRSAYESINELNNEONNYFEGYLHOEGGDCPIFKKRFK 1550
Db 2154 QOTKTIQNNPKD-----EIDDLIQELVYNNKESLKLPTIINKNDVTEPII-SRIDK 2205
Y 1551 LMGTSLAHSELSHKTAKINLSKVVDLIYDKENIDR--SNHRNPSDVLLDHAKFKIK 1607
Db 2206 VIN--LIKSEYNNNDVSNVAKKLE--EDANNIIRDLDTSHNMLND--LIQKNFKI- 2256
Qy 1608 FANGELIDPCAPNKHMKIMIONLOEIIIRNFRQPPVYN 1647
Db 2257 -----IDDLKNNKOEIEN-RNNIQTINREOEITQTEHVN 2289

RESULT 5

microtubule-associated protein MAP 1B - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
C:Accession: A56577
R:Zauner, W.; Kratz, J.; Staunton, J.; Felck, P.; Witche, G.
Eur. J. Cell Biol. 57, 66-74, 1992
A:Title: Identification of two distinct microtubule binding domains on recombinant rat
A:Reference number: A56577; MUID:92347374; PMID:1639092
A:Accession: A56577
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2364 <ZAU>
A:Cross-references: GB:X60550
A:Experimental source: brain
A:Note: nucleotide sequence not given; conceptual translation not complete
C:superfamily: microtubule-associated protein MAP1B

Query Match 4.0%; Score 342; DB 2; Length 2364;
Best Local Similarity 17.8%; Pred. No. 1.9e-06;
Matches 349; Conservative 313; Mismatches 695; Indels 608; Gaps 92;
Qy 2 NSTPKLLP-IDKSHLQD--OPOS-----SASIFNSPTKPLNF--PRTSKPSLIDPNSSS 53
Db 396 NSTQYNIIEGLEKTLHDLKPLATQKQKLTGOVSTPPKQYKIKLORADRSRESLKP--AT 453
Qy 54 DTYTSEDOOEKGEKKKDTAFQTSFDRNFDLNSIDIQOTIQHOQOOPQOQOOLSOQDNN 113
Db 454 KPLSSKSVYRKESKEAPEKTKASQVEKT-----PKVESKEKVIYKKDKPGK----- 499
Qy 114 LIDESFQTPMTSTLDLTKQNTVDKVNENHAPTYINTSPKNS-IMKATTPKASPKKAVF 172
Db 500 -----VESKPS-----VTEKEVPSKEEOSPVAEVAEAKATSKSPKVTKD 539
Qy 173 TVTNEIHHYPPNRYEEDOSO--OKEDSV-----EPPLIQHQWKPSPFNVSDE--DTN 223
Db 540 KYVKEIKITKPEKKEKKEKAKEDKTPKLKDEPKPEKKEKKEIKKEKEKEELK 599
Qy 224 ASVPPTPLHTTKPTFAOLLNKNNEVNSPEALJTMKILRENFSNLSLDEKYNLYLSPTN 283
Db 600 KEVKKETPLKDAK-----KEVKKDEKKEVKKKEKPEKKEIKKISKDIKSTPLSDTK 651

Qy 284 NNNS--KNVSDMDSHLQ-----NLQDASK-----AKTNENIIN-----LSF 317
Db 652 KPAALKPKYAKKEEPTKPEKPIAGKILDKGKVKIKKEGTTTEAATPANGTAVAAGA 711
Qy 318 ALKAPKNDIENPLNLTNDISLRSSGSS-----OSSLOSLRNDNR 358
Db 712 AASGPAKLE-----AERSLMSSPEDLTQDFELKAEIDVAADIKPQLELIDEEK 763
Qy 359 VLESVP-----GSPKKVNPGLSLNDGIKGFSDVEVESLP-----RDLSR 398
Db 764 LKTEBGEAVYIOKETEVKSGSAESPDEGITTEG--EGCEQEPFEELEPEVEKGVDDIEK 822
Qy 399 -----OKLETTKEHDAPEHNENNFIAKSTNTNKKGQLVSSDDHDS 440
Db 823 FEDEGAFGEESSEADYEERAE--TEBAEEDEEEDVNGSGASKHP-----TEDEIAK 876
Qy 441 FRSYNNHTQSILNLNSASQSLN-ALERQROTQEOBQTAAPREETSFDNITKY 499
Db 877 AEADV-HIKEKRESVASGDDRAEDMDALEK-----GEAQSEBEEBEEEDKADAREED 931
Qy 500 QEP-KSNLE-FYKVTIK-----EPVSATEIKAKKREFSRIILIRKNEDE 542
Db 932 HEPDKTEADYVAAVVDKAAEAGVTEDQYDFLCTPAKQPGVQSPSREPASSI-----HDE 986
Qy 543 IAEPAIHPKKEANEASHVEDTALLKALNDEESDTQNSTKMSIRHIID----- 594
Db 987 -----TLPGSSEEAATASOENRE-----DQPEEFTATSGYQSTIEISSEPTPMDEMS 1035
Qy 595 -----SDMKLESDNDGREDND-----DISRFKS-----DILNDVSQTSIDIG 634
Db 1036 TPRDVTMTDNNETSPSPQEFYNITKYESSLYSQEYKRYVASFNGLSGSGTDTDGR 1095
Qy 635 KYGNSSSEITTKTLAPRSDNDNDKNSKSLDEPANNSIQ-----QOLEVPTKEDDSIIA 690
Db 1096 DVNASAS-----TISPSSWEEDKFKSKALRDAYREEDTVKGAELDKVDYDERLSPA 1150
Qy 691 NSNNIAP-----PEELLTPVYE-----ANDYSSNDVTKTDAVSPFESISREH--E 736
Db 1151 KPSLSLSPSPSPLEKT-PLGERSVNSFLPNETKASAEBAATVAVSPGYQAAVEEHCAS 1209
Qy 737 TDSKPINFISIMHKOEKKHQTINIKVPTKOIYAS-----YQYKKEQES-----RYSQK 786
Db 1210 PEKKTLEVVS-----PSQSYGSAHTPTFYQSPTEKSKSHLPTVEYENA 1253
Qy 787 VKIPNALIQKKEKEVNVMSRRVYSPDMD-----LNVSQLPELSDSGFKDL 834
Db 1254 QAVPVSEFTEAKDEN--ERSSTISIP-MDEPVDPSEPIEKVLSPRLIGSESAVED- 1309
Qy 835 NFPAVYNNNNRRPSTPLSTKYNLSMIDNDPNVYEPPEKPSYAEIINARLSANKAAPNG 894
Db 1310 -FLSADDKALGRSESPFEKCKNGKQGSFQKESVSLDLSLQDKOEKRAK----- 1360
Qy 895 APPLPQROPSPSTRSNSNRKVRFRVPTFEIRTSALAPCDMYNDIFPD--FGAGSKPT 952
Db 1361 FLPIKEDFPEKKASDA-----EIMSSQALAA-----LDERKILGGGSGPT 1400
Qy 953 IKAEGMKTLPAMD-----KDDVKRILNAKKGVTOXEYINAKLVQDKPKKNSIVTDPED 1005
Db 1401 -----QVDVSQFGSFKEEDTK--MSISEGTVSQDK--SAPPVDE-----GAED 1437
Qy 1006 RYEELOQTASINNATIDSSISYGRPDSISTDMLPYLDELKRP-----PTALLSADR 1056
Db 1438 TYSHMEGVASVSTASVATSSF--PEPTTDDVPSLSLAEGVSPSTEVSDLSVSYQVPT 1495
Qy 1057 LFMEOGVHPLRSNSVLYVHGAGAGATNSMLPEPDFELINSAPRN--VSNNSDVAWALSQNA 1114
Db 1496 TQOETMSPSKED-----CPRPMGISPPDFSPKTAKSRTPVQDHRSESSMS----- 1542
Qy 1115 STISFNO-----LDNMFDP-----DQATIG----- 1133
Db 1543 --IEFGQESPEHSLAMDFFRSQSPDHPPTVGAGMLHITENGPTVEVDYSPSDIQDSSLSHKIP 1600

[illegible]

Db	2611	DAEAEVSKSDASRPSPVESVETAKDSTKGD1SPSPESVLEGPKDVEKSK--ESSRPSVS	2665
Qy	120	FOPTPTSTLDTLQKWPVDKVNENHAPTYINTSPNKSIMK-----KATPKASPKV	170
Db	2670	ASITGDSTKIVSRPASVESYKDEHDA-----ESRRESIAKVESYIDEAKSDSKSSSQ--	2724
Qy	171	AFTVTNPEIHHYPDNRVFEEDQSQOKEDSVAPPLIOHQMKPSPQFNYSDDETNASVPTP	230
Db	2725	-----DSOKDEKSTLAKEASRRESVESKSDAKSSRSPESYIAGEPVP	2771
Qy	231	-----PL-----HTTP-----TFAOQLNNENVESEPALUD--MKIKRENEN-----LS	270
Db	2772	RESKSPLOSKTDRPGSVESVYAEDEKSDQGRSRRSVAESYKADTKKQKQSOEASRPSS	2831
Qy	271	IDEKVNLYLSPNTNNNSKNYSMDSH--LONLDD-----ASKNKTENIMHLSFALKAP	322
Db	2832	VDE---LTKDDEKQOESRRQSTIGSHAMSTMDEDESPMKADSKRSPRESVASIKHE	2888
Qy	323	K-NDIENPLNSLTJNADISLRSSGSSQSL--QSLRDNRLVESYPSGPKKYNBGLSANDGI	380
Db	2889	NKDEESPL-----GSRDPSVAESITKSDITGKEKSPLEKSPKSVSPRESVSGI	2935
Qy	381	KGFQSE-----VSESILPROLSRKLET--TKRHDAPEHNNEFIDAKTNKQOLLYS	433
Db	2936	KOEKESRRESVAVSR--ESSKATYSAPPKESHSPRESYLSGLDDEGDKTTSRRVSAD	2994
Qy	434	S---DDHLDFSPRSYNHTEOSILTNLSASOSQ-----ISLNALEKOROTQEOQTOA	484
Db	2995	SLKDEKSLVSOEASRRESAEKSLKDAANASQETSRPESVTEBYVDGKSPVASKASRPA	3054
Qy	485	BPEETSTSDNIKVKOEKPSNLEFVKYTK--KEPVSAETELKAPREFSSRLIKNE--	540
Db	3055	SVAEAKMSADESEQRPESLPQSKAGSIKDEKSP--ASKDEAKESKEESRRESVAEQFP	3113
Qy	541	---DEIAPADIIHKKEMENSHNEODALLKALN-----DDEEDTQNTKMSI	589
Db	3114	LVSKEVSRPASVAESVDEAEKSEESPLSKASRSPASVAGSVKDEAEKSEESRRESV	3173
Qy	590	REFIDSDMK-----LEDSDNDREDNDDIRPEKSDILN-DVSQTDI---IGD	634
Db	3174	AEKSPLEPKASRPASVAESVKPDADKSKESRREGCAKSPASKASRPAVAESIKD	3233
Qy	635	KYGNSSSF-----ITTYTLAPRNSDNDKENSLEDPANNESLQOOLEVPTTKEDDS--	687
Db	3234	EAEEKSESRRESVAEKSPLEKASRPSTVASKVD-----EAEKSKESRSD	3282
Qy	688	ILANSNTIAPPEELTLRVANDYSFNRNDYTKFDYASSFEELSREHETOSKPLNTF--S	746
Db	3283	SVAEKSPILA--SKASRPAVAESVODEAEKSEESRRESVAKSPLAYKASRPAVAES	3341
Qy	747	IWHKOEKQKH-QIHKVPTQOIIAS-----YQYKNEOE-----	779
Db	3342	IKDEAKESKEISRRESVAEKSPASKASRPSTVAESVDEAEKSEESSRDSVAESPL	3401
Qy	780	-SRVTSDKVKIIPNAIQ--FKKFEVNVMSHRVYSPMDLANYSOFLPELSEDSQKDLNF	836
Db	3402	ASKASRPAVAESVODEAEKSEK-----ESRRESVAEKSPASKASRPAVAESVKD--D	3456
Qy	837	ANYSNNTNRPSF---TPLSTKANVLSINDNDPVVPEPPKSYA-----ELRNARLSAN	888
Db	3457	AEKSEESRRESVAEKSPASK-----EASRPAVAESVDEAEKSEESRR	3503
Qy	889	KAAPNOAP-PLPPOROPSSTRSNMKNRVSRFRPPEIRRTS-----SALACDWNNDIFD	943
Db	3504	ESVAEKSPLEKASRPTVAESVKODEAKSEK-----ESRRESVAEKSSLASKEF-----	3553
Qy	944	DEGASRPTIKAEQMKTLPSMDODVYKRIILNAKKVQOEVYINAKLVQCKPKKNSIYDP	1003
Db	3554	-----ASRPAVAESVKODEAKSEESRR-----ESVAEKSPASKASRPAVAESVKDE	3604
Qy	1004	EDRYEELQOQTASt-HNATIDDSIYGRPDSIS-----TDMLPYLSDE	1043
Db	3605	AEKKEVSRRESVAKSPLEKSPKASRPSTVAESVKDEADKSEESRREGAKSPKPLSME	3664


```

OY 1094 INSPARNNSNDVAISGMASTISFNOLMDPFOATIGKIOE-----QPAASKANVVR 1149
Db 1242 QSIORENDEKEKQLSEKDEKLQSIOQNLMNLNDENEDEKVKQFSEKDEKLQSIOQLNOLIK 1301
OY 1150 GDDDLASAPETPRTPRKESISSPKAKLSASPSPRIKGSPPVARIKKNGSIAIEPI 1209
Db 1302 QENQ-----EKQKQSEKDEKLQSIOQLNQLNDNQ-----IKKNELAKE-- 1343
OY 1210 PKATIKPKKSPFGNELSNHKVADGGISPSGSGEHQQHNPSWVSYSQTDAITVPDENK 1269
Db 1344 -EQLLKLODFDQDSOOLKOLEELSEKENOLOLKOENELNOLNOODOS-----NE 1395
OY 1270 DVONHREKOKNHNNHHNNHKKOTDIRGVVDDEIRDV-GLQERKGLFFRVLTGIKNIN 1328
Db 1396 ITIQLKDLLKQOOEOEDENNKE-----TEKIDIELDKQOEIDQSELSENKEIK 1448
1329 LPDINTHGKGRFTLTLDNGVCVTPEPYNMDDHNAIGKEFEFLTVDSEFIITLKASYEK 1388
1449 IQ-----TTQEFDQLSHN---RSKDQHILHQIQLEQLDKLQSDDD 1486
OY 1389 PRGTLEVETEKKVVKSRNBL-----SLRFSK-DITTTKFVPEVKDTMANKFAPD-- 1439
Db 1487 Q-----DHQFKKVIDERNYMLQOLEOSTLSNNQLDLLEKELPLELDSNEKOKTIDL 1541
OY 1440 GSFARCYIDLOOFEDQINGKASF-----DLNCFEMEWMTNGNOMPGRGPYKIA 1490
Db 1542 SNISMQISLQNDKLIKISERNNSITLESRTIQLSLDEKKNLIKLOQKQKQO---- 1596
OY 1491 QLEVMKLVVPRSDPREI-----LPTSIRSAYESINELNNEONNYPFGYLHOEGDCPIFK 1546
Db 1597 QQPPTASSPSSPSFLSTPTPKQRPNQIEDRLVNEIVRNQDDLIRKN-----KT 1649
OY 1547 RFFKLMGTSLAHSEISHKTRAKINKLKYVDLIIYDKEN 1585
Db 1650 KEYKLENGDYIVNSII---YRLSLDDNDSDLIAQEYEN 1685

RESULT 8
TI18674
hypochemical protein T04F3.1 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C.Accession: TI18674; T24464
R.White, S.
submitted to the EMBL Data Library, June 1996
Reference number: Z19004
Accession: TI18674
Status: preliminary; translated from GB/EMBL/DDBJ
A.Molecule type: DNA
A.Residues: 1-3147 <WIL>
A.Cross-references: EMBL:Z74026; PIDN:CAA98419.1; GSPDB:GN00023; CESP:T04F3.1
A.Experimental source: clone B0240
R.Kershaw, J.
submitted to the EMBL Data Library, May 1996
Reference number: Z19894
Accession: T24464
A.Status: preliminary; translated from GB/EMBL/DDBJ
A.Molecule type: DNA
A.Residues: 1-3147 <WI2>
A.Cross-references: EMBL:Z72513; PIDN:CAA96672.1; GSPDB:GN00023; CESP:T04F3.1
A.Experimental source: clone T04F3
C.Genetics:
A.Gene: CESP:T04F3.1
A.Map position: 5
A.Introns: 338/3; 417/3; 1957/2; 2358/3; 2400/L; 2529/2; 2585/L; 2700/3; 2744/L; 2877/3

Query Match 3.8%; Score 331.5; DB 2; Length 3147;
Best Local Similarity 18.8%; Pred. No. 7.7e-06;
Matches 349; Conservative 283; Mismatches 666; Indels 561; Gaps 82;

OY 2 NSTPS-KLLPLDIKSHILOPOSSASIFENSPTPLNPNFRNSKPSLDPN----- 50
Db 422 SLTSPRHRRAPD---TEL-----DLEN--PKPFSPANSKRPPPPRRRRHPSPAS 469
```

QY	51	-----	SSDPYT	57
Db	470	VAHSOSFVDEPSQHFVTTTIDRNOYTPVTTTTTKRESGRLHENSINOLLSSDMNV	529	
QY	58	SE-----ODQKEGEEKKDAFQTSFD--NFPLDINSIDQOTIHQOQOOPQOQOOLSQT	110	
Db	530	AQMMNDASSEMENNVS-----SKSSLSLTSSTSAKASVARIISVDELTKREKRSQAVPLPS	589	
QY	111	DNNLIDFSOTPM-----TSLDLTKON--PYUDKVA--ENHAPTYINTSPKNSIMKK	160	
Db	590	DVSDIDIFTALTTPKREVOSEKRFVITTRQORADVADIDFEKDRPTP-----	638	
QY	161	ATPKASPKVAFVTYNBEIHHYDNRVEEDSOQKEDSEPERPLIDHQKDESOQFNS--	218	
Db	639	APPSGKRKSATATNOGEITLLELDEVEKHELDSSKVSTSTIN---LNDESMETRNNDSD	695	
QY	219	--DEDTN-----ASVPPTPLHTTKPTF---AOLKNNEVNSEPALTDMKLRE-	264	
Db	696	SFDEEVPNRNOLVLEIPESEPRYTSTATIOLESSVAGENSE--NKRP--VISMKSSEI	752	
QY	265	-----NFSULSDEKVNLYLSPTNNNNSKNVSDMSHLQNLQ	301	
Db	753	AKKEKDAQRSQFVIITPSKEHIDDESIVSMDVFN---TPHDCRPDPIDA-----	801	
QY	302	DASKNKTN-----NIHNSFLAKARKNDIENPLSLTNADISL-----	340	
Db	802	DSKHTSDSDSREVSTYITINDV--FPTEBPK---LVAKDCIEAEERIGRIKQF	854	
QY	341	-RSSGSSQSSLOSLRNNRYLES-----VPSGPKVNVGSLNDKIGCF	383	
Db	855	ERTTGEQGISNSEPTDEMDKDRHTSAVSIDLKVFQGTAKKPE-----ND--EF	906	
QY	384	SDEVESLPLPKDSRLKLETTKEHDA--PEINNENFIIDAKST-----NTNKQLLVSSD	436	
Db	907	DEKIRGIAFERSKQKEVQORSGVATSHSGKHIDESISMVDVENTSQ--KYKSD	963	
QY	437	HLDSDFSYN--HTSQSLTNLNSASOSQISLNLKOROTOEDQOQAAPEEFTSFSN	495	
Db	964	KLSPERPERVEVSTATMNLDNISASGIATR--EENTNVLEEERIOKRVPEFKTTEN	1021	
QY	496	IKVQOEPSNLEFY--KVTIKKEPVASATEIKAPREESSRLIRKNDSE-----IAEPAD	549	
Db	1022	LEIQEVYLTKEEVNDSDVKEHRSAAVNLDEKVLFIHSGSKKAKKNDKIRGIAFEFT	1081	
QY	550	HPKKEANEANSHVEDTQALLKALNDEESQOTONSTKMSIRHIDSQMLKEDSNDGRED	609	
Db	1082	KQKEEAQSTVIE-----TSQSN--RIFEESISMDDVFNNSLHN	1120	
QY	610	NDDISRFKSDILNDVQSOTDIIGDKYGNSSSEITTKTLAPRSDNNDKENSKL----	664	
Db	1121	ESQVSEITTEADPEBDLVLTSTTFINVTEEKIDDDVYKT-----DSNVEEKEQVRLRD	1174	
QY	665	--EDPANNESSLOQLLEVPHTKEDDSILANSSNIAPPEELTLPVYEANDYSFENDYTKTFD	722	
Db	1175	EFKRPTPEONLOKQEFELTKKEEESYV-----KMEKRTSAVSIDLKVED	1218	
QY	723	AVSPEESLSHEHTDSKPINFTSIMHKQEKQKHQIHVPTKQIIASVQOQKNOQESV	782	
Db	1219	-QSSKEFTVS--NETDEKIRGIAFERSKQKEVORSGAE---TSHSGKHIDESNI	1271	
QY	783	TSDDVKATPNAIOFKKFEVVMVMSRRVSPMDLANSQFLPELSDSGKDLNFANSNN	842	
Db	1272	SMDVFN--NNSQKXKXDEKLSSPRYEVEPVSTATMNN--LDNITTFASG-----ATREEN	1322	
QY	843	TN-----RPRSTPLSTKKNVLSINDDPNVVLEPEPYSYATIFRNARLSANKAAP	892	
Db	1323	TDVLEEERIOKRVEEF-----KKTENLEIQEVYLTKEEVDNSOVKCHRTSAVINDL	1377	
QY	893	NOAPRPLPQORPQSTRSNKKRVSRFPVFFELRPTS-----SALACDMYN-----	939	
Db	1378	DVF-----IQRSSKHPEBNDEDEKIRGIAFERKQKEAORSVAIETSSQKHIEDP	1433	


```

Db 947 TNNDEKHTSKINDENITIDKO-KGLNNPVDLNRKIKNE-----TKILEKDSHNS 1000
Oy 1061 QEVHPL-----RSNSVLVHPGAGATNSMLPE--PPELINSBARVSN-----NS 1105
Db 1001 KQNNLMKKKTPEPTNNKGISSTSSSTKFKDCGITEK-NKLNLSLKTYCIRKKMAY 1059
Oy 1106 DNVAISGNASTISFNQDMNFDQATIGOKIOEPASKSANTYRGDDGLASAPETPRP 1165
Db 1060 DSIRINDKA-----DLKDKKTSFNDINR--AAKGMNFKKRD-----VP 1097
Oy 1166 TKKESISSPRAKLSASPKSPKIKGSPYVTK-----KNGSIAGIEPIRK 1211
Db 1098 MNNNVDTKKGK-----RVFNRYTLNNNNYNNYIRSNKNVKNKGMVGIKIVL 1146
Oy 1212 ATKPKKSPQGNIEISNKKVRDGGISPSGSEHQOHNPASVVSQYTDATSTVPDENKV 1271
Db 1147 LKEKOK-----SLHPEGEAKKLLSNYNNDYVLEKDFKDI 1183
Oy 1272 QHKPREKOKNNNNNNNNNNKOKTDIPGVVDELPDVGLOERKGLFEVLAGIKNI----- 1327
Db 1184 INEEMEKYK-----NNKKYKIK--SNSIPPI-----IKKIERKS 1216
Oy 1328 NLPDINTHNGRTLLDNCVH-----CV-----TTPRYNMDHNVAIGKEPEL----- 1370
Db 1217 NDNNDNINNNNDNINSNNNNNNNDKCLPLSKEDRVHLKNNNIYNNNTMPEKQNSCD 1276
Oy 1371 TVADSLFETLTKASYEKRGTLYEVEKVKVSRNRSLRSGSKDI----- 1418
Db 1277 NNTTSLKNNMIINSSEKANSNSQNNNEKMSYFEMLANEKKKKKLTIEKDEKKKGEL 1336
Oy 1419 -----TTTKFVPTVEKVDTWANKFAPDGFARCYIDLOOFDOITKASOFDNCENWE 1472
Db 1337 EKSEHANDSKKDEQLDADKPTMLQPLSF-----NLQONERKY--EQGDFYIDKY----- 1385
Oy 1473 TMSNGNQP-----KRGPRYKI-----AQLEYKMLYVRSRPRELLPSISAYSINE 1521
Db 1386 PFGNGTGTGVFAIKIIEKKENKRVALKVMAKDTMSLNIEROVLKELI--IQASLKHNI 1442
Oy 1522 LNNQNNPEEGYLOHGGCGPIFKKRFKFLM----- 1552
Db 1443 L-ELIAYED-----KTRFLILELANGSVRNKKOKKQPLNEEVALYVF 1488
Oy 1553 -----GTSLSAHSEISHKTRAKINLSKVVDLIYVDKENIDRSNHRNSDVLLDHAKEIKF 1608
Db 1489 QIADALSTYHNHNIIHRODKPDI-----LIHYSNEHLNKKIKYK-IGVIKIADFGSCQL 1542
Oy 1609 ANGE-----LIDFCAP 1619
Db 1543 KKKRKRSTFCGTIDYMPR 1561

```

RESULT 11

UC6009

Surface-located membrane protein Imp3 precursor - Mycoplasma hominis

C:Species: Mycoplasma hominis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: Jc6009

R:Label: Jc6009

U: Bacteriol. 178, 2775-2784, 1996

A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis Imp gene system

A:Reference number: Jc6009; MUID:96213016; PMID:8631664

A:Accession: Jc6009

A:Molecule type: DNA

A:Residues: 1-1302 <LAD>

A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CA64858.1; PID:g1197336

C:Genetics:

A:Gene: Imp3

C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology

C:Keywords: duplication; membrane protein

F:1-2/Domain: signal sequence status predicted <SIS>

F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>

F:957-992/Domain: tetratricopeptide repeat homology <TTL>

```

F:993-1026/Domain: tetratricopeptide repeat homology <T2>
F:1089-1120/Domain: tetratricopeptide repeat homology <T3>
F:1154-1190/Domain: tetratricopeptide repeat homology <T4>

```

Query Match 3.8%; Score 324; DB 1; Length 1302;

Best Local Similarity 20.6%; Pred. No. 4.9e-06;

Matches 281; Conservative 221; Mismatches 463; Indels 398; Gaps 70;

```

Oy 9 LPDK-----HSHLOLPQSSASIFN-----SPKPLN 37
Db 150 LPDKTKTKNNENENAKKELLKINARELO-----SKTFNEKKQELKRYLLEDTREVD 204
Oy 38 FPRTPNSPLDPSSSDTYTSEDOEGKEEKKDTAFOTSFDRNFDLNSIDIOQTIOHQ 97
Db 205 FTK-EQGVLEFN-INETSIEDIKKIIIEVAKTSLS-----KILNKK 248
Oy 98 QQQPQQ-----QQQLSOTDNNLIDFSFQTPMTSTDLTKONPTVDKVENENHAPTYINTSP 153
Db 249 QOELQEFENIKKDIQFIPTKILNDAKYOSIKQALD--KIN-SLNGINKSTYKEIKAGQ 305
Oy 154 NKSIMKATPRKASPKK-----VAFTVNP-----EIHHPDNRV 187
Db 306 NALL--KAKEEAGELEKEKLDGONIKDTLKTETINNAKEPKKLLIDNDQKIVDKSLNDEL 363
Oy 188 EEDQSQOKE-DSEVEPPLIQHWKDPDSOFNYSDEDTNASVPRPPLHTTKPTFAQLLNK 246
Db 364 SKAQGLSLMDKESME-----SANDL-----LNTKILIEKELLK- 397
Oy 247 NEVNSPEEA-LTDMKLRKRENFSLIDEKVN-----LYLSPNN-----NNSKNVSD 292
Db 398 --FNOEKAKFNFLEQTRKNIENFLTDEVKNNPNVATLVKGLDINAKDAKSVYSSNKKSD 455
Oy 293 M-----DSHLODASKKNNTNENIHLSPALKAPKDIEPLNSL----- 333
Db 456 IIAANEALQALADA--NNAKDOVDANKSITKQLNALDIAKATLLPOLNDNDELYKAK 513
Oy 334 --TNADISLRSSGSSQSSLSLDRNDNRVLESVSPKVPGLSLNDGKIGFSEVESL 391
Db 514 ESINAEITNANKRVNQ-----NDNASMQSAKS-----SLDDKYVTKIQNLTE-- 555
Oy 392 LPDLSRDKLETTKEHDAPENHNNFT--DAKSTNTKQGL--LVSSDDHLDSFDRSYNH 447
Db 556 ---FMKDKDAKFELEQTRKIDNFLTDDVKNPNVATLVKGLDIAKADKKSVTKSSNK 611
Oy 448 TEOSTLNLSASQSSISLNALEKQROTOEQOEAEPREESFSDNITKVKQEPSNLE 507
Db 612 SE--II-----AANDELK-QALDKARVAKDQID-----BANNSIKQLSDSTT 651
Oy 508 FVKVTIKKEPVSAETIKAPREFSSRLRIKNEDETAEPADIRPKKENANSHEVDTDAL 567
Db 652 NANOGLNKLVDSDKOQAKKTELISOETQSASQELNINPTSMQSAKS-----LDAK 703
Oy 568 LKKALNDEEDSTYQNSTKWSIRFHIDSDWKLDESDNGDREDNDISREPKSDILNDVSO 627
Db 704 VTEI-----TKLETFNKKDKVYKPEL-EKTRKIDDEFINTKTN----- 742
Oy 628 TSDIIDKYGNSSETITTKLAPRSDNNKESKSLD--PANNESLOQLLEVPHTKED 665
Db 743 -----PNSTLSIELSK-----RDSKNSVTSSNKSDETANTE-LKQALAKANDK- 789
Oy 686 DSLIANSSNIA--PPEELTLPVEANDYSFND-VTKTFPAYSFPESLSRHEHETSKPI 742
Db 790 ---AQADNLAKSTKEQLN-----NSISSANTLLAKLTDKMTIQQA----- 827
Oy 743 NFSTIMHKOEKQKHQ-IHKVPTKQIIASYQYKKNDEQSVYSDKVKIPNAIOFKKFEY 801
Db 828 -----KTELEKEVQKANOAVASNNTASMQSAKSIDAKVTEITKKLET---FNKDKDV 877
Oy 802 NWSRRVSPDMD--LNVSQFLPE-----LSESDGFLD--NFNYSNNTNRRPSFPLST 854
Db 878 KFKELQTRKIDDEFINTKTNPDYSTLSIELSKDSKNSITNSNKSIDETANTELQ 937
Oy 855 KNLNSINDND--PNVVEPPEPKSYAEIRNARRLANKAAPNQAAPLPQPQPSSTRSN 912

```

Db 938 ALAKANTKQADADNARSTKEQLNKSISANTLLAKLT-----DKDNTIQOAK 985
QY 913 KRVSFRVPTFEIRRTSALAPCDMYNIDFDEFGAGSKPTIK-AGCKTLPMDQ-DVK 970
Db 986 TELER-----EVQANQAVAS-----NNTASQSAKSSISADKAVETITTKLEFFNKDKDK 1035
QY 971 -RIINAKKGVTODEYIN-----AKLVQDKPKNSIV-----TPDEYEELO 1011
Db 1036 FRELEQTRKDI--DEFINTKNTPNVSTLISELTSKROSKNSITSSNKSQDIETANTLEK 1093
QY 1012 QTASHTNNTIOTSSICGRDSTIDMLPYLSDELKPPAL-----LSADRLFMEDE 1062
Db 1094 Q--ALAKANTDKA--QADNARSTKEQLNKSISANTLLAKLTDKDNTIQOAKTELEKE 1148
QY 1063 V---HPLRSNSVLVHPAGATNASSML-----PEPDF-----ELIN 1095
Db 1149 VQKQAVASNTNTVSMQSAKSSLDTKVETITTKLEFFNKDKAKNELKKTGQIQOEFIN 1208
QY 1096 SPARNVSNNSDVA-----ISGNASTISFNOLDNPFDDQATIGQIQOAPASKANTVR 1149
Db 1209 T-NKNPNVSELISOLTSKROSKNSVTNNSKSDIE-----TANTELKQ--ALAKANTDK 1260
QY 1150 GDDDLASAPETPRPTTKE---STSSKPAKLSASPRKSPK 1189
Db 1261 AOADNLARS-----TKEQLNKSISANTLLAKLTDKDNTIQ 1296

RESULT 12
S45781
probable calcium-binding protein YBL047c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBL0520
C:Species: Saccharomyces cerevisiae

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Apr-2002

C:Accession: S45781; S50284; S45782; S37339; S42458
R:Goffeau, A.; Joniaux, J.L.; Purrelle, B.; Skala, J.; de Wergifosse, P.; van Dyck, L.
submitted to the Protein Sequence Database, August 1994

A:Reference number: S45745
A:Accession: S45781

A:Molecule type: DNA
A:Residues: 1-961 <DEF>

A:Cross-references: EMBL:Z35808; GSPDB:GNO0002; MIPS:YBL047C
A:Experimental source: strain S288C

R:de Wergifosse, P.; Jacques, B.; Joniaux, J.L.; Purrelle, B.; Skala, J.; Goffeau, A.
Yeast 10, 1489-1496, 1994

Title: The sequence of a 22.4 kb DNA fragment from the left arm of yeast chromosome II
binding protein.
Reference number: S50284; MUID:95176707; PMID:7871888

A:Accession: S50284
A:Molecule type: DNA

A:Residues: 1-961 <DEF>

A:Cross-references: EMBL:X78214

A:Experimental source: strain S288C

R:Dubois, E.; El Bakoury, M.; Glandsdorff, N.; Messenguy, F.; Plerard, A.; Scherens, B.;
submitted to the Protein Sequence Database, August 1994

A:Reference number: S45782
A:Accession: S45782

A:Molecule type: DNA
A:Residues: 579-1381 <DUB>

A:Cross-references: EMBL:Z35808; GSPDB:GNO0002; MIPS:YBL047C

R:Schierens, B.; el Bakoury, M.; Vlerendeels, F.; Dubois, E.; Messenguy, F.
Yeast 9, 1355-1371, 1993

A>Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of ye
A:Reference number: S39824; MUID:94205266; PMID:8154187

A:Accession: S39841
A:Molecule type: DNA

A:Residues: 579-1381 <SCH>
A:Cross-references: EMBL:Z23261; NID:G313733; PIDN:CAA80797.1; PID:G313748
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SCD-EDL1; MIPS:YBL047C
A:Cross-references: SCD:S0000143

A:Map position: 2L
C:Superfamily: yeast probable calcium-binding protein YBL047c; calmodulin repeat homo
C:Keywords: calcium binding; EF hand; transmembrane protein
F:167-199/Domain: calmodulin repeat homology <EF1>
F:560-576/Domain: transmembrane #status predicted <TM>

Query Match 3.7%; Score 320.5; DB 1; Length 1381;
Best Local Similarity 19.5%; Pred. No. 7.4e-06;
Matches 229; Conservative 183; Mismatches 456; Indels 309; Gaps 48;

QY 3 STPSKLLP--IDKSHILOLOPOSSASIFNSPTKLANPRTNRSK-----LDPNSSSD 54
Db 378 SAPQIAIPSRASKPSLODMPHOVSAPAVNTQPTVQVLPONSNNGLDALLANFSFSP 437
QY 55 TYSQDEQEKKEKKDPAFCPSFDPNFDLSDIDIOQTIGHQOQPOOQOOLSTQNNL 114
Db 438 SPT-----KAQYVQNTNNSFSYDNN-NGQATL--QOQOPOQPPPLTHSSGL 483
QY 115 IDESFQTPMTSTL--DLTKQNPVQKVENHAPTYINTSPKSIIMKATPK--ASPKV 170
Db 484 -KKF---TP-TSNFQSIIEKEPEEQOLRESSDTF-----SAQPPVPKHAASPVR 531
QY 171 AFTVNPPEIHHPDNRVEEDDSQOKEDSVPRLLQHQMKDPSQFN-----YSDEDTNA 224
Db 532 TASTLLPOVPNPSVSPAGATSAATGAAGAALGASAFSSNNAPKQDLPA 591
QY 225 SVPTPLHTTKPTFAQLLNKNEVNSEPALTDK-----LKRENFSLDEKVMLY 278
Db 592 DGEASQSLSNATTEMANISNOVNSLSKQASITNDKSKRATQELKAVTEMKNSIQKLN-N 650
QY 279 LSPTNNNSKNVSDMSHLQNLQDASKKNTNENHNSLPALKAPKNDIENPLNSTL----- 334
Db 651 LRSTHDQVKKQTEQLEAVLQV-----NKENETLAQQLAVSANYHAESKINELTDLQ 705
QY 335 -----NADI--SLRSSGSSQSLSLRDNRLVLESPSPKVVN-PGSLND----- 378
Db 706 ESQTNNAELKEQITLNSMTASLQSLNEKQOQVQKQSVNDVNSKQLELQVYANLQK 765
QY 379 GIKGFSDEVESILPRDLSRDLKLETTKEHDAPENHNEFTDAKSTN-TNQQLVSSDH 437
Db 766 EIDGGEKISVYLTKQKELNDYQKTEVEHQALQKQYDLSKDKDLDLREKQLEERRNQ 825
QY 438 LDFPRSRNHTBQSILNLNSASQOISLNALEKQROBOBOQTOAEPBEETSNDIK 497
Db 826 IEEQNLVHQHYSKLQEMFDDLSQKASFEKADQELK-----ERNIEAVNVR 873
QY 498 VKQEKSNL-----EFKVTIKREPVSATEIKAKRFPSSILIKNDEIAPADTIP 551
Db 874 ELSEKQMLAMQLPEDAKDIATK---SASNTDTTKEATSR-----GVNHE 917
QY 552 KKENE-ANSHVEDTDLALKALNDEESDTQNSTKMSIRFIID-----SDMKLEDSN 603
Db 918 DTVSKFVETVNSMLNVRVAKDEKTERTESDV-----FDRDVLPGSQSDSENANTN 972
QY 604 DGDREDNDISFEKSDILNDVQTSIDILGDKYGSSEITTKTLAPPRSDNNKENS 663
Db 973 NGTQSGNETAN---PNLTETLSDRFDDLDMEYGIPRSQSLTSSV----- 1013
QY 664 LEDPANN--ESLQQLVPHTKEDDSILANSN-----IAPBELTLPVVEANDYSS 713
Db 1014 -----ANNAPOSRDVLEPTELEERTINNTANRDNNTNLSHIPEWEATPATASTDVL 1069
QY 714 FNDVYKTDVAVSSFEESLSREHETSKPINFISWHKQOKKHQIHVYPTKQIIASVQ 773
Db 1070 -NETIEVEDSGTTKRAANSN---DGEVSST---QESPKTSAQPAKTI----- 1112
QY 774 YKNEQSRVTSQKIPNAIOFKKFEVNVMSRVVSPMDMLDVSQFLPELSEDSGFKD 833
Db 1113 ---NEE-----PPIDELHIDESSSDSDDERED 1139
QY 834 LNFANYSNNTNRPSEFTPLSTKNVSLNIDNPVPEPEPKSYAETIRAKRLSANKAPN 893
Db 1140 TR-----EIPSAIVYKTLQTPYN-----AQPTSSLEIHT-----E 1168

Db 1780 YNNLKKYCDMEHEDSLMLIKHLVLLKDEKENIKNNEGSESDNLLID-----KNMKS 1835
Qy 1405 RNRSLRLEFGSKDITTTKFEVPTVDKMTANKFAPDGSFARCYIDLQOFEDQITGKASOPD 1464
Db 1836 DQENE--NEKDHNNIEYI-ENLSEHLRHKY-----DIEMIPIDIVGNIRLKK 1881
Qy 1465 LNCFNEMET-----MSNGQPM-----KRGKPK-IAOLEVKKLY-----1499
Db 1882 LFYINNIKNNYSYLKNNFPPKYKRGKAKGMLIMIKRKHEDYDNLMKKNGNEKGT 1941
Qy 1500 -----PRSDREILPTIRSAYESINELNNEQNNYFEGYVLEHOG-----1539
Db 1942 NNNKKNNKEDNDYINNN-----NNNNNNNNNDNNINNGYKRMQNNINPNLYETIKNN 1996
Qy 1540 -----DOPIFKRFKLMG-----TSLAHSEI--1562
Db 1997 FNLHYRWSLFKYCPVSYINDNKIIEGKKFCVYIKNKIYMLNNDYIEFLKHPKFA 2056
Db 1563 SHKTRAKILSLVY-----DLIYV-----KE 1584
Qy 2057 SKKVESVCSFLVYFPNMDIQKIIYLETYEKLHIDIGYFKQYFPLSHQDKBOHKN 2116
Qy 1585 NIDRSNHRNFSVLLLD-HAFKI-KFANGELID-----FCAPNK--HEMKIWNLOE 1633
Db 2117 NKKSNNDODVONKIEDAHIFEIHRITEGKTISPFLYVYICIMHKTIQFNIYI--IKE 2174
Qy 1634 IYRNRFRQPVNMLQOOQ 1654
Db 2175 II-KSIILKLNHFNLLMKK 2194

RESULT 14
ORMSPL
microtubule-associated protein MAP1B - mouse
N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text change 01-Sep-2000
C:Accession: S07549; S44387; A33645
R:Noble, M.; Lewis, S.A.; Cowan, N.J.
J:Cell Biol. 109, 3367-3376, 1989
A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A:Reference number: A33645; MUID:90094539; PMID:2480963
A:Accession: S07549
A:Molecule type: mRNA
A:Residues: 1-2464 <NO>
C:Cross-references: EMBL:X51396; MID:952999; PIDN:CA35761.1; PID:953000
Manchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Ch. Biochem. Biophys. 310, 428-432, 1994
A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A:Reference number: S44387; MUID:94234720; PMID:8179328
A:Accession: S44387
A:Status: preliminary
A:Molecule type: protein
A:Residues: 653-663, 'IC' <SAN>
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein; tandem repeat
F:589-786/Domain: microtubule binding #status experimental <MTB>
F:589-592, 639-642, 649-652, 655-658, 663-668, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69
R-K-E/D-X)
F:1861-2064/Region: 17-residue repeats
F:91, 116, 351, 888, 1124, 1153, 1168, 1208, 1662, 1877, 1918, 2003, 2030, 2054, 2083/Binding site: p
F:147, 969, 1336, 1562, 1563, 1702, 1708, 1990, 2057, 2063, 2419/Binding site: phosphate (Thr) (cc
F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 3.6%; Score 314; DB 1; Length 2464;
Best Local Similarity 17.7%; Pred. No. 3e-05;
Matches 341; Conservative 319; Mismatches 736; Indels 534; Gaps 92;

Qy 2 NSTPSKLLP-IDKSHLQI--OPOSSSASIFNS-PTKPLN---FPRTNSKPSLDPNSSS 53
Db 492 NSTQVNLLEGLEKTLKHLFLKQPLATOKDLTGQVTPPVKQVKLORADSRSLKP--AT 549
Qy 54 DTYTSDQDQEKKEKKKTAFTQTSFDRNFDLNSIDIOQTIOHQOQFQOQOQLSQTDMN 113

Db 550 KPVASKSVKREKKEETPEVTKTSQVEKT-----PKVESKEKVLVKKDKPVKTESKPSVTEK 605
Qy 114 LIDEFSPMTSTL---DLTKNP--TVDKYNNENAPYI-----NTSPNKIMKK--AT 162
Db 606 EYSSKEQSPVAEVAEKQATESKPKVTKQYVKEIKLEKKEEKKKEVKKEDKT 665
Qy 163 PKASPKVAFVTYNPEIHYPDNRYEEDOSO-QKEDSVPEPL-----IQH 207
Db 666 PLKDEKPRKEVKKEL-----KKEIKKEERKELKEVKKETPLDKAKKKEVKKKEVKK 721
Qy 208 QMKDP--SQNYDEDTNASVPPTPLHTTKPTF--AQLLKNNNEVNSEPALTDKK-----260
Db 722 EKEPKKEIKISKDIKS---TPQSDTKPKALPKVAKKEESTKKEPLAAGKLKDKG 778
Qy 261 ---LKRNFNSN-----LSDEKVNLYLSTPNNNNSKN 289
Db 779 VKVYKKEGTTAAATVGTAAATTAAYVAAGIAASGPVKELEAERSLMSFEDLTKEE 838
Qy 290 -----VSDMDSLQNLQDASKNKTNE--NIHNLSPALKAPKNDIENPLNSTLNDI 338
Db 839 ELKAEIDVAKDIKLPQLELIEDEKLETKQPEAVYIQETEVSGSASPDGITTTE- 897
Qy 339 SLRSSSSQSSLSQSLRDNKRVLESVPGSKKYNP-----GLSLNDIGKGSDE---VVESL 391
Db 898 ---GEQCEQ-----TPLELPEVKQGV---DDIEKFEDEGAGFEES 934
Qy 392 LPRLSDRLKLETTKEHADPEHNNENFIDAKSTNTKQGLVSSDDHLDSFDSYNNTEOS 451
Db 935 ETGDY-EEKAE-TEEAPEEEDGE---DNASSASKSHSTEDDESKAADVHLKEKRS 989
Qy 452 IINLNASQSOQISLNALEKOROTOEOQTOAEEETSFSDNIKVKQEP-KSNLE--F 508
Db 990 VVSGCDRAEEDMD--DYLEKGEAQSEE---EEBEEKADAREBEGEPDKTEADQV 1042
Qy 509 VVVTIKKEVSAFTE-----IKAKRRESSILIKNDETAFAADLHPKKE 554
Db 1043 MAVADAAAGVTEBOYGYLTSAKQPGIQSPSREPASSI-----HDE---TLPGSE 1092
Qy 555 NEANSHEVEDTALLKALNDEESPTONS-----TKMSIFHIDSPK 598
Db 1093 SEATMSDEEN-----REDPEEFATSGYGTQSTIEISSEPTIMDESTRDVMSET 1144
Qy 599 LEDSDGREDNDISREKS-----DILNDVQSOTDIIGDKYGNSSSETT 644
Db 1145 NMEETESQSEFVNITTKYESSLYSOEYKPAVASFNGLSGSKTADTGKOVNASAS---1201
Qy 645 TKTLPAPPSDNDKNSKSLERPANNESIQ-----QOLEPHTKEDDSILANSNAP---697
Db 1202 --TISPSSMEEDKFSKSLRDVAYCSEKELKASALDIDKVDLSRSPAKSPSLSPSP 1259
Qy 698 -PEELTLPVVEANDYSF--NDVTKTFDAYS-----SFEESLSREH--ETDSKPINFISI 747
Db 1260 SPIEKT-PLGESSVNFSLTPNEIKVASBEAASVSPGYQVAVEHCHASPEKTELEVVS-1317
Qy 748 WHKQEKKKHQTINHYPTKQIIAS-----YQQYKNQOES---RYTSDKVIKPIAIOFKK 797
Db 1318 -----PQSVTGSAGHTPYQSPDTOKESHLLPTEVSEMAQAVPVSEFSE 1362
Qy 799 FKEVNVMSKRVVSPMDI-----LAVSQFLPELSDSQFKDLNFANYSNNNTNR 845
Db 1363 AKDEN--EKASLSP-MDEVVPDSPEVKEVLSPLRSPPLIGSESPYED--FLSADSKVIG 1417
Qy 846 PRSFTPLSTKVNLSNIDNDPNVPEPEPKSVAEINARLSANKAQAPOAPLP-----899
Db 1418 RSESSEPECKKQKQF-----PDRESVPYSDLTSTGLVODKQKEESTGTGTPKEDFG 1468
Qy 900 PQROPSTRSNSNKRVRSPRFPTEIRTSALACDQWYNDIFDDFGAGSKPTIAEGMK 959
Db 1469 PEKRTSDVETMSQAL-----ALDERKLGDSVSPQTQIDVQFGSF-----1509
Qy 960 TLPSMDKDDVRIILAKKGVQDQETINAKLVQDKKKNSIYTPDPRDKLEEDQOTASTIMA 1019

Db 1510 -----KEDTK--MSISEGTVSDK--SATPVDEGV-----AEDTSHMEGVASVSTA 1551
 QY 1020 TIDSSYGRDSTIDMLPLVLSDELKRP-----PTALLSDRLFMQEVHPLRSN- 1069
 Db 1552 SVATSSF--PEPTDDVSPGLAEVSGPHSTEVDDSLSVSVQPTPTTQETEMSPSKSEEC 1609
 QY 1070 ---SVVHPGAGATNSSMP-----EPPDFLIN 1095
 Db 1610 PRMMSISPPDFSPKTKAKSRFPVQDHRSEQSSMSLEFGQSEPHSFANDFSRQSDHTTLG 1669
 QY 1096 SPARNVSNNSD-----NVAISGNASTI-----SFNOLDNMFDOATIGOKIOBPA 1141
 Db 1670 ASVLHITENGPTEVDYSPCDIODSSLSHKIPTEEPSTYQ-DNDLSLISVSO-VEASPS 1177
 QY 1142 SKANANVRGDDGLASPELTPRTPKKESIS-----SK 1174
 Db 1728 TSSAHT-----PSQIASPLDIEDTLDVPPREMSLYASLASEKVOSLEGELSP 1776
 QY 1175 PAKLSASPRKSPKIKIGSP-----VRVKKNGS---TAGIEPIPKATHKP---KKSFOGN 1223
 Db 1777 KSDISPLTPRESS-PLXSPFSDSTSAKETAHAHQSSSPIDAIATAEYGFRRSMFLD 1835
 QY 1224 EISNHKVRGDISPSS-GSEHQONPSMVSVPQYTDATSTVDENKDVQHKPREKQK 1282
 Db 1836 TMOHHLALNBDLTSSVEKDSGKTPGDFNVAYOKPENAGSPDE-EDYDESQEKTIPT 1894
 QY 1283 HHHHHHHHKKOTDIPGVVD-----EIPDVG-----LQERGLTF 1319
 Db 1895 HDVVRYYEKTERT-IKSPCDSGSYETIEKTKTPEDGGYCEITEKTRTPREGGYSY 1953
 QY 1320 RVLGKININLPDIN-----THKGRFTL-TLDNGVHCVTPEVNMDDHNAIGKEFELT- 1371
 Db 1954 EI-SEKTRTRPEVSGTYEKTERTSRLLDISNCYDDEGCHLIGCQSYETIEKTRITS 2012
 QY 1372 --VADSLFTLLKASTEKRGTLVEYTERKVVAKSRNRLSRLGSKD-IITTKFVPTEV 1428
 Db 2013 FPESSESYSYETSTRTKSPPTSAVCYETMEKIKTPQASYSYETSPRCYCTTEKKSPSSEA 2072
 QY 1429 KOTWANKFADGSPARCYIDLOQFEQITKASQF-----DLNCFNEMETMSNGQPKR- 1483
 Db 2073 KQ-----DVIDCLVSSCEFKAPKTELSPFINPNDLEWAGEEPPESEKPLQOS 2122
 QY 1484 -GPKYKIAOLEVLMYVPRSDPR-----ELPSTIRSAVESINEINNEQNNYFEGYLH 1535
 Db 2123 GGAP-----PRSGKGQGGQGCDEPTPTSVSESAPSOOTSD-----VP 2159
 QY 1536 QEGGDCPTEFKRRFKLMGTSLAHSEISHKTRAK-INLSKVVDLIYDKENI---DRSNH 1591
 Db 2160 PETEECP-----SITADANIDSEDESETIPTDKTVYTKHMDPPAPWQDSRPS 2207
 QY 1592 RNFSDVLLLD 1601
 Db 2208 PRHPDVSMVD 2217

RESULT 15

T34418
 hypothetical protein f12f3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34418

R:Fulton, B.; Wohlmann, P.
 submitted to the EMBL Data Library, July 1998

A:Description: The sequence of C. elegans cosmid f12f3.

A:Reference number: Z21521

A:Accession: T34418
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-3448 <FNU>

A:Cross-references: EMBL:U800022; PDB:AAC25865.1; GSPDB:GN00023; CESP:f12f3.3
 A:Experimental source: Strain Bristol N2; Clone f12f3

C:Genetics:
 A:Gene: CESP:f12f3.3

A:Map position: 5
 A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 3.68; Score 313.5; DB 2; Length 3448;

Best Local Similarity 18.44; Pred. No. 5e-05;

Matches 344; Conservative 282; Mismatches 649; Indels 593; Gaps 80;

QY 43 SKPSLDNNSSDITYTSFQDQEKKEKKDPNAPFQSPFRNPDLDNSIDIQITQHQOQOPO 102
 Db 330 SEESRDDKSYD-----EVDOSTYLEKKDDGDKSKRPT-----KKKLLKKETPE 376
 QY 103 -----OQOQLSOTDNNLLIDEFSF-----OTPMSTP 127
 Db 377 SEQVTAAPREQQKISEVDQVSVATEVGAKKKPAPAEKPTDLAKKSKSKSKSEPEAST 436
 QY 128 LD-LTKONPVDKYNENHA-----PYINTS--PN 154
 Db 437 EEKSTTEKPTNDKTSKSKSAEKKTVKPKREVTKGLEAKKPYEDKDKASOPSSKESSPPT 496
 QY 155 KSIKKKAPPKA--SPKKVAFVTNPETIHHYPDNVE-----EEDSQQKEDSVEP----- 202
 Db 497 DGKKKKOIPRALFTPELISRFQDPTMSETNITTTIRGREGSADAKTPLVEPLSASVS 556
 QY 203 ----PLIOHQKDPSCQFNY-----SDEDTNAYVPPTPLHTTKPTFAQLLNKNNEVNSE- 252
 Db 557 MKVFTLVESA-KEKAEEFPKRSETPPDKSKRKGELP-----AKKSEKKEVTAEK 607
 QY 253 --PALTDMLKRENFNSLDEKVNLYLSTNNNSKNVS-----DMDSHLQNLQD 302
 Db 608 QSTFALIESKKVEDESKISEQ-----PSDKNKSEVGVKPAKAPETKKVSEIEE 660
 QY 303 ASKKKTENIHNLSFALKAPKNDIENPLNSLTNADISLRSSGSSQSS-LOSLRNDNYLE 361
 Db 661 VPKKKTITK-----KTEKSD-----SISOKSNVLRADDDKSKSD 696
 QY 362 SVSPSPKRVNPGSLNDGKIGFSDVEVESLPRDLSDKLETTTEHAPENHNENFTDAK 421
 Db 697 DVTQSKK-----TTEQTKVATQSKLEK-----AAQTKQIE-----TEVYVDK 737
 QY 422 STNTNKKQLVSSDDHLDSPFRSYNHTEOSLNLSNASQSO--ISLNALEKQOTQBOE 479
 Db 738 S--KKKVLKKTKEKSDSFISQSETPPVETPKPASEKQKTAEVAKKQKEVEVDNL 793
 QY 480 QTOAAPPEEETSFSDNJKVKQOEPSNLEFVTKKEKPEVSATEIKAPREBSSTLRKN 539
 Db 794 KREAEVAAKTA--DEKLKLEAEN-----IKK-----TAEVAAKKQ-KEKDEQKL 838
 QY 540 EDEIAEPADIIHPKKENANSHE--DTDALLK-KALNDEESDTTONSTKMSIRFHDS 595
 Db 839 ETEVYSSKSAAEKLELEKQAKIKKAADAVAKQKELNEKKLEAKKSAADKL----- 892
 QY 596 DKLIEDSNDGRENDDISREKSDILNDVQSDTIDGKIGNSSSETITTTTLAPPSDN 655
 Db 893 --KLEE-----ESAASKKRYSEESKVEEKKRTAGETKYVQVESEPTSK 935
 QY 656 NDKENSKSLDEPPANNESLQOQLEVPHTKEDDSILANSSNIAPPELTLPVVEANDYSFN 715
 Db 936 TIDRKDVGATPRA-DETPKKKIITKKTEKSS----- 966
 QY 716 DVTKTFAYSSFEELSREHETDSKPINFISIMHKOEKKQKHQIHKVPTKOILASVQYK 775
 Db 967 -----SISOKSATDSEKVS-----KQKQODE-----PTKRAVSQTQVVT 1000
 QY 776 NEQES--RVNSDKVKIPNALQFKFKKQVNVMSRRVYSPMDLDLNVSGFLELSDSQSF- 831
 Db 1001 EADSKSRQKETDEKLDAELAATKQEDADSK-----LDAQKRIKKVSEDDAAR 1051
 QY 832 --KDLNFANYSNNTNRPSPLSTKNVLSINDPNVVEPPEPSYAEINARLT----- 885
 Db 1052 KEKELN-----DKLKESEIATKKASA--DKLKEEQAKKAAEVAEAAKKQKQKD 1100
 QY 886 -----SANKAAPNQAAPLPQOR-----PSTRSNSKRVYSRFRVP 921

```

Db 1101 EQLKLDTEAASKAAAEKLEKQAOIKKAAGADAVKKOKLEDEKKNLEANKKSAAGKL- 1159
QY 922 TFEIRRTSSALAPCDWYNDIFDDPGAGSKPTTIAEGMKTLPSMDKDDVKKILNAKKGVQ 981
Db 1160 -----KIEESAASAKSKOTVEEQAKLDAQTKAKTAKEQTKLEKDEKST 1201
QY 982 DEYINAKLYDOKPKKNSIYTDPEDRYEELQOTASINATIDSSITGRPDSIS----- 1033
Db 1202 KESESEKYDEKPKKKVLLKKTESDSSISOKSETSKTVESA--GPSESETOKVADAAR 1259
QY 1034 -----TDMLEPYLSDLEKPPALSLADRLPMEQVHPLRSNYLVHPGAGATNSMLPEP 1089
Db 1260 KQKETDEKOKLEAEI-----TAKKSAD-----EKSKLEAEKLLKKAAYEAANKQKEKDE 1309
QY 1090 DEFLINSPA-----RNVSNNSDNVAISGNASTISFNOLD 1123
Db 1310 QLKLDTEAASKAAAEKLEKQSHIKKAAYDAVKKOKLEKOKRLESEAAATKKADAER 1369
1124 MNFDDQ-----ATIG-----OKIOQPAKSAANTYRGDDGLASA-----PETPRTPK-- 1167
Db 1370 LKLEEQKKAAEIALIEIQOEKLAQEOS--RLEDEAKKSAKOKLESETKSKQTEEA 1426
QY 1168 -KESISSKPAKL-----SSASPRKSPK-----IGSPVRIKKNGSINGIEP 1208
Db 1427 PKESVDEKPKKKVLLKKTESDSSISOKSKSAKSTYDAAEETLESDFNLVEKK-TVQKVEQ 1485
QY 1209 IP-----KATHKPKKSPQNEISNH-----KYRDGGISPSGSEHQHNPMSVSPQYT 1258
Db 1486 SPDESTATIKRDPQOKTEEISQODDGEKTTTDDGPKPKPESEATPKKRVYKKKTQKS 1545
QY 1259 DAT-----STVPDEK--DYQHKPREKQKOKHHHHHHKQKTDIPGVVDEIPDYGLQ 1312
Db 1546 DSVASDASLADVSKLSDVEEKPKKKVL-----KKKTEKSDSVISSETSSVD-- 1591
QY 1313 ERGKLEFRVIGIKINILPDINTHKGRFTLLDNGVHCVTTPPEYMDMDHVAIGKEPELTV 1372
Db 1592 -----TIKESVSEIP-----TEKAQMLLN-----RFS---- 1615
QY 1373 ADSLEFILTLKASYEKPRGTLEVEYTEKV--VKSNNRLSRLFGSKDITTTKFPV---T 1426
Db 1616 -----TDSAVSESEPKNAHKODTEKTTDDMTTRKSSAIFSDPOSISSTSEGRRR 1667
QY 1427 EYKDTWANKFAPD-----GSFACYIDLOQFEDQITGKASQFDLNCFNEMETMS----- 1475
Db 1668 RRRTGFAASKFASDTLALRGDNVEIEAELLAEEDTVTKVNGKADLNSRCHESHTEFRT 1727
1476 ---NGNOPMKRGKP-----YKIAOLEVKNL-YVPR-----S 1502
Db 1728 LIIDEVEPTDSMEITATGTSHTTILKVEELPVDFVKYLPKRTSGKEQEVYISVTLN 1787
QY 1503 DPREI-----LPTSIRSAVESINELNNEQNNYFEGYLIHQEGGDCPIFKKRFPKLM- 1552
Db 1788 HPIDISKVYVWLKDGKPLEINKDY-SIDTVGCSYSLTLRAKAYEDSGK-----YKVV 1838
QY 1553 -GTSLLAHSEISHKTAKINLSKYVDLIYVDKENIDRSNHRNFSVDYLLLDHAFKIKFA-- 1609
Db 1839 DGVDSTHLSIOGKPYLK-NVSETPKPIYVDDK-----QFSLVAYADSNPEASFSMT 1890
QY 1610 -NGELIDF 1616
Db 1891 VDGKDLDF 1898

```

Search completed: March 17, 2003, 12:26:55
Job time : 115.814 secs

FT	CARBOHYD	638	638	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	670	670	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	691	691	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	817	817	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	838	838	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	841	841	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1018	1018	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1082	1082	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1100	1100	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1103	1103	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1113	1113	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1200	1200	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1571	1571	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1593	1593	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	DOMAIN	95	106	POLY-GLN.	(GLCNAC. .)	(POTENTIAL)
FT	DOMAIN	283	286	POLY-ASN.	(GLCNAC. .)	(POTENTIAL)
FT	DOMAIN	1283	1292	POLY-HIS.	(GLCNAC. .)	(POTENTIAL)
FT	DOMAIN	1651	1660	POLY-GLN.	(GLCNAC. .)	(POTENTIAL)
FT	SEQUENCE	1664	AA; 187859	MM; BA2EFD0DC8196790	CRC64;	

Query Match	100.0%	Score 8631;	DB 1;	Length 1664;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1664; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	MNSUPSKLPLDKHSHLOLOQOSSSASTFNSPTRPLMFPRRNSKPSLDPNASSSPTTYSBQ	60
Dp	1	MNSUPSKLPLDKHSHLOLOQOSSSASTFNSPTRPLMFPRRNSKPSLDPNASSSPTTYSBQ	60
QY	61	DOEGKEEKKDTAOTCFSDFRNFEDLJNLSIDLOQTIOHOOQOOPQOQOOLSQTDNNILIDEFSF	120
Dp	61	DOEGKEEKKDTAOTCFSDFRNFEDLJNLSIDLOQTIOHOOQOOPQOQOOLSQTDNNILIDEFSF	120
QY	121	QTPMTSTLDLTKNPPTYDKVNHNAPTYINTSPKKSIMKKATPKASPKVAFTVTNBEIH	180
Dp	121	QTPMTSTLDLTKNPPTYDKVNHNAPTYINTSPKKSIMKKATPKASPKVAFTVTNBEIH	180
QY	181	HYPNRVAEEEOOQOKEDESVPEPLIOHOMQDPQOFNYSDEPTNASVPTPEPLHTTKTPFA	240
Dp	181	HYPNRVAEEEOOQOKEDESVPEPLIOHOMQDPQOFNYSDEPTNASVPTPEPLHTTKTPFA	240
QY	241	OLLNKKNNEVSEPALTDMLKRENFENLSLDEEYVNLISPTNNNNSKNVSDMOHLONL	300
Dp	241	OLLNKKNNEVSEPALTDMLKRENFENLSLDEEYVNLISPTNNNNSKNVSDMOHLONL	300
QY	301	QDASKNKNTNENIHNLSFALKAPKNDIENPLNLSLTNADISLRSSGSSQSSLOSLRNDNRVL	360
Dp	301	QDASKNKNTNENIHNLSFALKAPKNDIENPLNLSLTNADISLRSSGSSQSSLOSLRNDNRVL	360
QY	361	ESVGSPPKVPNGLSLNDGICGSEDEVEYESSLPDLSRDKLETTKEHDAEBHNNEFIDA	420
Dp	361	ESVGSPPKVPNGLSLNDGICGSEDEVEYESSLPDLSRDKLETTKEHDAEBHNNEFIDA	420
QY	421	KSTYTNKGQLLVSSDDHLDSFDRSYNHTEOSILNULNLSASOSQISLWALKKOROTOEOQ	480
Dp	421	KSTYTNKGQLLVSSDDHLDSFDRSYNHTEOSILNULNLSASOSQISLWALKKOROTOEOQ	480
QY	481	TOAAPEEETSFSONIIVKOEPPKSNLEFVYVITTKKEPVSAETELKAPKREFFSRILRIKNE	540
Dp	481	TOAAPEEETSFSONIIVKOEPPKSNLEFVYVITTKKEPVSAETELKAPKREFFSRILRIKNE	540
QY	541	DEIAEPDPIHKKRENEANSIHYEDTDALLKALNDEESOTTONSTKMSIRPHIDSDMKLE	600
Dp	541	DEIAEPDPIHKKRENEANSIHYEDTDALLKALNDEESOTTONSTKMSIRPHIDSDMKLE	600
QY	601	DSNGDREDNDIDISRFKESDILNVSQTSPIIGDKYNSSESEITTKTLAPRSDNNDEXN	660
Dp	601	DSNGDREDNDIDISRFKESDILNVSQTSPIIGDKYNSSESEITTKTLAPRSDNNDEXN	660
QY	661	SKSLEDPANNESSLOOQLEVPHTKEDDSILANSSNIAPPEELITLVPVEANDYSSFFNDVTKT	720
Dp	661	SKSLEDPANNESSLOOQLEVPHTKEDDSILANSSNIAPPEELITLVPVEANDYSSFFNDVTKT	720
QY	721	FDVAYSSEESISRGHEFTDSKPIFINEISIMWQEOKKKHQIHKVPTKQIIASYSQYKNEOES	780
Dp	721	FDVAYSSEESISRGHEFTDSKPIFINEISIMWQEOKKKHQIHKVPTKQIIASYSQYKNEOES	780

Db	721	FAVYSFEESLSRHEHETDSKINFLSIHMOEKKKQIHKVPRKQIILASTYQOYKKEDES	780
QY	781	RVTSQKVKIPNAIOFKKFEVNVMSRRVVSFDMDLNVSOFLPELSEDSGFKDLNFANTS	840
Db	781	RVTSQKVKIPNAIOFKKFEVNVMSRRVVSFDMDLNVSOFLPELSEDSGFKDLNFANTS	840
QY	841	NNTNPRSFPTLSTKKNVLSNIDNDPNVYPEREPEKSVAEINARLSNKAAPRQAPRLPR	900
Db	841	NNTNPRSFPTLSTKKNVLSNIDNDPNVYPEREPEKSVAEINARLSNKAAPRQAPRLPR	900
QY	901	QOPFSSTKSNKRVSRFRVPTFELIRRTSSALAPCDAYNDIPDFGAGSKPTIKAGSKMT	960
Db	901	QOPFSSTKSNKRVSRFRVPTFELIRRTSSALAPCDAYNDIPDFGAGSKPTIKAGSKMT	960
QY	961	LPSPMDKDVKRIILNAKKGVTODEYINAKLVQKPKNSIVTDEDRYEELQOYASIHMAT	1021
Db	961	LPSPMDKDVKRIILNAKKGVTODEYINAKLVQKPKNSIVTDEDRYEELQOYASIHMAT	1021
QY	1021	IDDSITYGRPDISTPDMPLRYSDELKCRPTLLASMDRLFMQOEYHPRJRSNVULNHPRAGA	1081
Db	1021	IDDSITYGRPDISTPDMPLRYSDELKCRPTLLASMDRLFMQOEYHPRJRSNVULNHPRAGA	1081
QY	1081	TNNSMLRPDEDELINSPARVNSNNSDNVAISGNASTISFNQOLMNFDDQATIGQKIOEOP	1141
Db	1081	TNNSMLRPDEDELINSPARVNSNNSDNVAISGNASTISFNQOLMNFDDQATIGQKIOEOP	1141
QY	1141	ASKSANVYRGDDGLASAPRRPTKKESSSKPRALSSAPRSKSPRIKGSVYRIYIKN	1201
Db	1141	ASKSANVYRGDDGLASAPRRPTKKESSSKPRALSSAPRSKSPRIKGSVYRIYIKN	1201
QY	1201	GSIAIEPIRPATKPKPSFQGNISINHKVRDGISFSSGSEHQNNPMSVSPQYDTA	1261
Db	1201	GSIAIEPIRPATKPKPSFQGNISINHKVRDGISFSSGSEHQNNPMSVSPQYDTA	1261
QY	1261	TSTYVDEKKDYOKRPREKOKOHHNNHHNNHOKDTPIGVUDEGTRDVGLOERGLKFR	1321
Db	1261	TSTYVDEKKDYOKRPREKOKOHHNNHHNNHOKDTPIGVUDEGTRDVGLOERGLKFR	1321
QY	1321	VLGJKNINLPIPIINHKGFTLLDNGVCHSTPBYNMNDHNVALIGKEFELTVADSLEFL	1381
Db	1321	VLGJKNINLPIPIINHKGFTLLDNGVCHSTPBYNMNDHNVALIGKEFELTVADSLEFL	1381
QY	1381	TLKASYEKPRCTLVEVEYEKKVYKSRNLSIRLFSKSDIITTTKVPURPEVAKDTJANKFPARG	1441
Db	1381	TLKASYEKPRCTLVEVEYEKKVYKSRNLSIRLFSKSDIITTTKVPURPEVAKDTJANKFPARG	1441
QY	1441	SFARCYIDLQOFEQITGKASQOEDLNCFNEMETMSNGNCPMKRGKPYKTAOLEVUKMLVP	1501
Db	1441	SFARCYIDLQOFEQITGKASQOEDLNCFNEMETMSNGNCPMKRGKPYKTAOLEVUKMLVP	1501
QY	1501	RSDPEILPTIRSAYESINELNNEOONNYEGYLHOBGGGCPFEKKRPFKLMGTSLIAMS	1561
Db	1501	RSDPEILPTIRSAYESINELNNEOONNYEGYLHOBGGGCPFEKKRPFKLMGTSLIAMS	1561
QY	1561	EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVDLLDHAFAIKTFANGELIDFGCAPN	1621
Db	1561	EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVDLLDHAFAIKTFANGELIDFGCAPN	1621
QY	1621	KHEKKIWTIONLEIITNNRRRORPQVWYLMIOQOOOQOOOQSSQ	1681
Db	1621	KHEKKIWTIONLEIITNNRRRORPQVWYLMIOQOOOQOOOQSSQ	1681

RESULT 2		
ID	BUD4_YEAST	PRT; 1447 AA.
AC	P47136.	STANDARD;
DI	01-FEB-1996 (Rel. 33, Created)	
DI	01-OCT-1996 (Rel. 34, Last sequence update)	
DI	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	Bud site selection protein BUD4.	
GN	BUD4 OR YJR092W OR J1905.	


```

Dd 1117 GNIRLNKEPELPIDPKGAETSSASSERSTYKCVITLCKEKEPRHELVETIVDPVPGK 1176
Qy 1404 SNRLSRLEGSKDITTTTKFVPEYK-DTWANKFAPDGSFARCYIDL-QOEFQDTGKAS 1461
Dd 1177 S-----FFGKTKYKFEKKYQKKRQDEMDYLFADQDSFACELIENEFKNAFNIS 1230
Qy 1462 QPDLNCFEMWETMS--NGNQ--PMKCKPYKIAOLEVKMLVPRSPDRELLPTPSIRSA 1515
Dd 1231 HHMYMKRQSRVADKINGSKRLYELPRKAPHKVASLDVEAFELERTSAFEQFPKQFSLV 1290
Qy 1516 YESINELNNEONNVEGYLHOEGDCP-IFKKRFFKMGTSLLASHSEIHKTRAKINLSK 1574
Dd 1291 NKIVSKYLQONIVKEGYLQDGGDLCKIENRFFKLHGSQLSGYHEISRKAKIDINLKL 1350
Qy 1575 VVDLIYVDKENIDRSNHNFSVDLLDHAFKIKFANGELIDFCAPKHEMKI-WIQLNQE 1633
Dd 1351 VTKVLARNEDIDQADNGGQNFPTDWFVFNCFQFLVFDGGERITFNACNSDEKSDWYNKQ 1410
Qy 1634 IIRRRFRQRPVNLMLQOQQOQQOQQSS 1662
Dd 1411 VVELNVF-HQPWVKYCKEKLAEERKRTT 1438

RESULT 3
MAPB-RAT STANDARD: PRT: 2459 AA.
ID MAPB-RAT STANDARD: PRT: 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [contains: MAP1
light chain LC1].
CN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Testis;
RX MEDLINE=96257242; PubMed=666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
Gene 172:307-308(1996).
[2]
RC SEQUENCE OF 96-2459 FROM N.A.; DOMAIN, AND INDUCTION.
RC STRAIN-Sprague-Dawley; TISSUE-Brain, and Gli1al tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Katz J., Staunton J., Felick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
RT Eur. J. Cell Biol. 57:66-74(1992).
[3]
RN [1]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE-Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rientz A., Gremingloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Litaier U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
nervous system that is immunologically related to microtubule-
associated protein 5.";
RT EMBO J. 8:2879-2888(1989).
[4]
RN [1]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Notthias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -1- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes

```

```

CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -1- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -1- INDUCTION: By nerve growth factor.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: TO MAP1A.
CC -1- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

```

```

DR EMBL: U52950; AAB17068.1; -.
DR EMBL: X60370; CAC16162.1; -.
DR EMBL: X16623; CAA34620.1; ALT-SEQ.
DR PIR: S06017; S06017.
DR InterPro: IPR000102; MAP1B-neuraxin.
DR Pfam: PF00414; MAP1B-neuraxin; 10.
DR PROSITE: PS00230; MAP1B-NEURAXIN; 8.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2459
FT REPEAT 1869 1885 MAP1B 1.
FT REPEAT 1886 1902 MAP1B 2.
FT REPEAT 1903 1919 MAP1B 3.
FT REPEAT 1920 1936 MAP1B 4.
FT REPEAT 1937 1953 MAP1B 5.
FT REPEAT 1954 1970 MAP1B 6.
FT REPEAT 1988 2004 MAP1B 7.
FT REPEAT 2005 2021 MAP1B 8.
FT REPEAT 2022 2038 MAP1B 9.
FT REPEAT 2039 2055 MAP1B 10.
FT DOMAIN 559 1035 GLU-RICH.
FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
FT DOMAIN 2224 2312 LYS-RICH.
FT CONFLICT 127 127 M -> V (IN REF. 1).
FT CONFLICT 140 140 T -> S (IN REF. 1).
FT CONFLICT 2112 2112 R -> K (IN REF. 3).
FT CONFLICT 2169 2169 L -> I (IN REF. 3).
SO SEQUENCE 2459 AA; 269497 MW; 2E3F6872DDBBBA2 CRC64;

```

```

Query Match 4.0%; Score 342; DB 1; Length 2459;
Best Local Similarity 17.8%; Pred. No. 3,2e-06;
Matches 349; Conservative 313; Mismatches 65; Indels 608; Gaps 92;

```

```

Qy 2 NSTSPKLLP-IDKSHQLD--QPOSS---SASIFNSPTKPLNF-PRTNKSPSLDPNSSS 53
Dd 491 NSTGVNILEGKELKHLDFLKQPLATQKDLGQVSTPVPVKVKKLRQDSRESLKP--AT 548
Qy 54 DTVTSQDQKNGKEKKKTAQTSTDRNFQDLSIDIQTTQHQOQOQPOQOQOQSLQDNN 113

```

Db 549 KPLSSKVSKESEKEAPEATKASQVEKT-----PKVSEKKEVIYKKDKPGK----- 594
Qy 114 LIDERSFOQPMSTLDLTQONPTVQKVNENHAPTYINTSPNKS-INKKATPKASPKVAF 172
Db 595-----VSKPS--VTEKEVPSKEQSPVAEVAEKATEKSPKVTYKD 634
Qy 173 TYTNPEIHNYPNRWEEEOQO--OKEDSV-----EPPLIOHMKPQSOQFNYSDE--DTN 223
Db 635 KVVKEIKTKPEEKKEEKKEKAKEDKEDTPLKDKPKKEKKEKKEKKEKELK 694
Qy 224 ASVPRTPLHTTKPTFAOLLKNNEVNSPEALTDKMLRENFNSLSDKEVNLSPTN 283
Db 695 KEVKKETPLDKA-----KEVKDEKKEVKEKEKPKKEIKKISKDKIKKSTPLSDRK 746
Qy 284 NNNS--KNVSDMSHQ-----NQDASK-----KTNNTNIN-----LSF 317
Db 747 KPAALKPKYAKKEEPKKEPIAAGKLKDKGVKVIKEKGTTEAATAVGTAAVAAGV 806
Qy 318 ALKAPKNDIENPLNTNDISLRSSGS-----OSSLOSLKNDNR 358
Db 807 AASGPAKLE-----AERSLMSSPEDLTDFEELKAELIDVARDIKROLELIEDEK 858
Qy 359 VLESVP-----GSPKVPNGLSLNDGIGKFSDEVESLIP-----RDLNR 398
Db 859 LKETEPEGAAYIOKETEVSQSAESPDEGITTEG--EGECEOTPEELEPEVEKQGVDDIEK 917
Qy 399-----DKLETTKHDARPEHNENFIDAKSTNTKKGOLLVSDHLDLS 440
Db 918 FEDEGAFESSEBAGDYEEKAE--TEEAPEPEDEGEDNVSGASAKHSP-----FEDEIAK 971
Qy 441 FDRSYHTQOSILNLNSASOSQISLN-ALEKOROTQEOFOAARPEETSQDNIKYK 499
Db 972 AENDV-HIKEKESVASGDDRAEDMDALEK-----GEAPOSEBEEBEEKADAREED 1026
Qy 500 QEB-KSNLE-FYKVITIK-----EPVSAETIKAPKEPSSRIILRIKNEDE 542
Db 1027 HEEDKTEADYVAVVDDKAAGVTEQDQVFLGTPAKOPVOQSPREPASSI-----HDE 1081
Qy 543 IAEPAIHKKEMEASHVEDTALLKALNDEESDTQNSKMSRRIID----- 594
Db 1082-----TLPGSESEATASDEENRE-----DOPEETATSGYOSTLEISSEPTPMDEMS 1130
Qy 595-----SDMKLEDSNDGREDND-----DISRFKES-----LILNDVQSOTDILIGD 634
Db 1131 TPBDVMTDETNNEETESPQGEFVNITKYESSLXSOEYKRVVASPGLSGSCTATADGR 1190
Qy 635 KICNGSSSEITTKTLAPPRSDNDKENSLSLEDPANNEILO-----QOLEVPHTEEDSILA 690
Db 1191 DYNASAS-----TISPPSSMEEDKFSKALRDAYRPEETDVKTGAELIDKVDSEDLSPA 1245
Qy 691 NSGNINAP-----PEELTLPVYE-----ANDYSSFNVDYKTFPDANSSPEESLSREH--E 736
Db 1246 KSPSLSPSPSPSPEKT-PLGERSVNFSLPNEIKASAGEATVAVSPGVTAQAAVEECAS 1304
Qy 737 TDSKPINFISIMHOKOKKHQIHKVPTKOILAS-----YQOYKNEOS-----RVTSK 786
Db 1305 PEKTELEVVS-----PSQSVTGSAGHTPRYQSPTEKSHLPLEVTEMA 1348
Qy 787 VKIPNALIQFKKEKENVMSRRVVSPPMD-----LNVSOFLPELSDSGFKDL 834
Db 1349 QAVPVAFEFTEAKDEN--ERSSISIP-MDEPVPDSESPIEKVLSPLSPLIGSESAVED- 1404
Qy 835 NFANYSNNNTNRRSFPTPLSKNYLSNIDNDPNVYPERPEKSYAELIRARLSANKAAPQ 894
Db 1405 -FLUSADKALGRSESPFECKNKQGFSDSESVDLTSDLYDCKEERKAG----- 1455
Qy 895 APLEPQROPSSSTRNSNKNVSRFVPTFEIRRTSSALAPCDMYNDIFD--FGAGSKPT 952
Db 1456 FIIKEDFSPEKKASDA-----EIMSQSALA-----LDRKRLCGDGSPT 1495
Qy 953 IKAEGMKTLPSMD-----KDDVYKRLNKKKGVTODEYINAKLVQDKPKKNSIVTDPED 1005
Db 1496-----QVDSQFGSFKEDTK--MSISEGTVSQK--SAMPVDE-----GAED 1532

Qy 1006 REBELQOATASIHATIDSSISYGRPDSISTDMLPYLSDELKP-----PTALLSADR 1056
Db 1533 TYSHMEGVASVSTASVATSSF--PEPTTDVSLSLAEVGSPPSTEVDOSLSVQOTPL 1590
Qy 1057 LFMEOEVHPLRSNVLVHRCAGATNSMLPEPDFELINSAPRN--VSNNSDNVAISGNA 1114
Db 1591 TPQETEMSPSKKE-----CPRMWSISPPDFSPKIAKSTPRVQDHRHSGSSMS--- 1637
Qy 1115 STISFNO-----LDNMF-----DOATIG----- 1133
Db 1638 --TEFQOSEPHSLAMDFFSROSPDHPYVAGMLITENGTEVDYSPSDIODSSLHKIP 1695
Qy 1134-----OKIOEPASKANTVRDODDGLASAPENPRPTKESSIS-- 1172
Db 1696 PTPEPSTIODNDLSELLSVQVAPSPSTSAHT-----PSOIASPLOEDTLSDV 1744
Qy 1173-----SKPAKLSASAPRSPIKIGSPVRYLKKNGSIAGIE 1207
Db 1745 VPPRDMSLVASLASEKVQSLGEKLSPKSDISPLTPRES-----SPYSPGFSDSISGAK 1799
Qy 1208-----PIKATHKP--KKSFGNGLSNKRVKDGISPPS--GSEHOQHNPBW 1251
Db 1800 ESTAAAYQTSPPSIDAAAEYGFRRSSMLFTQMHNIALSRDLTSSVEKDNCKGKTPDGE 1859
Qy 1252 SVPSQYTDATSTVPRDNKQVQHKPRKOKKNNNNNNNNKKTDIRGVVD----- 1304
Db 1860 NVAQKPESTESSPDE--EDVDYSHKETTQAHNVGSGYIETKRT--IKSPCDSGSYETI 1917
Qy 1305-----EIPRVG-----LOERKLEFRVLGIKINIMLPIN-----THKGRFTL- 1341
Db 1918 EKTTPKEDGQSCETLEKTTTRPEEGGYUEL--SEKTTTRPEVSGYTEKTRSRRLD 1976
Qy 1342 TLONGVCHVTTRYNMDDNVAIGKPELTVADSLERLITLAKSYEKPRSTIYEVTEKY 1401
Db 1977 DISNGYDTEDEGHTLDGCSYETTEKIT-----SPRESESYE-----TTTKT 2022
Qy 1402 VKSRNRLSRFGSKDITITTKFVPTVEKVDWANKFAPD--GSFAR-----CYIDLOQFE 1453
Db 2023 TRSPDTSAYETEMETIKTTKTPQASTSYETSQDKYTERKSPBARQDVCLVSSCEFK 2082
Qy 1454 DQITGKASQF-----DLNCFNEMETMSNGNQPMKR--GPKYIAOLEVKMLVPRSDPR- 1505
Db 2083 HPKTELSPPRINPNLEMFAGEPTESERPRLTQSGAP-----PPSGGKOQ 2129
Qy 1506-----ELPRTSISAVESIENLNNEQNNNYFEGILHOGGCGCPFKKFKFLMTSLLAHS 1560
Db 2130 GROCDETPRPSVSSEAPSOQTDSD-----VPRPTECP-----SITADA 2167
Qy 1561 EISHKTRAK--INLSKVYDLIYVKENT--DRSNHRNFSVLLLD 1601
Db 2168 NLDSDESETTPYDKTYTTHKMDPRAPMODRSPRPHRYSNVD 2212

RESULT 4
YBET YEAST
ID YBET YEAST STANDARD; PRT; 1381 AA.
AC P34216;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 150.8 kDa protein in SEC17-QCRL intergenic region.
GN YBL047C OR YBL0520 OR YBL0501.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN (1)
RP SEQUENCE OF 1-961 FROM N.A.
RC STRAIN=5286;
RX MEDLINE=95176707; PubMed=7871888;
RA de Wergifosse P., Jacques B., Joniaux J.-L., Purnelle B., Skala J.,
Goffeau A.,

```

RT RT "The sequence of a 22.4 kb DNA fragment from the left arm of yeast
RT chromosome II reveals homologues to bacterial proline synthetase and
RT murine alpha-adaptin, as well as a new permease and a DNA-binding
RT protein."
RT RL Yeast 10:1489-1496(1994).
RT (2)
RN RN
RN RN SEQUENCE OF 579-1381 FROM N.A.
RP RP STRAIN-S288C:
RC RC MEDLINE=94205266; PubMed=8154187;
RX RX Scherens B., el Bakoury M., Vierendeels F., Dubois E., Messenguy F.;
RA RA "Sequencing and functional analysis of a 32,560 bp segment on the
RT left arm of yeast chromosome II. Identification of 26 open reading
RT frames, including the RPL1 and SEC17 genes."
RT RL Yeast 9:1355-1371(1993).
CC CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC CC -1- SIMILARITY: CONTAINS 3 EH DOMAINS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL: Z35808; CAA84867.1; -.
DR DR EMBL: X78214; CAA55048.1; -.
DR DR EMBL: Z23261; CAA80797.1; -.
DR DR PIR: S45781; S45781.
DR DR SGD: S0000143; YBL047C.
DR DR InterPro: IPR000261; EPS15_repeat.
DR DR InterPro: IPR000449; UBA_domain.
DR DR Pfam: PF00036; efhand; 2.
DR DR Pfam: PF00627; UBA; 1.
DR DR SMART: SM00054; Efh; 2.
DR DR SMART: SM00027; EH; 3.
DR DR SMART: SM00165; UBA; 1.
DR DR PROSITE: PS00031; EH; 3.
KW KW Hypothetical protein; Repeat.
FT FT DOMAIN 14 113 EH 1.
FT FT DOMAIN 135 227 EH 2.
FT FT DOMAIN 277 366 EH 3.
SQ SQ SEQUENCE 1381 AA; 150783 MW; 626fde261DCBA7D99 CRC64;

Query Match 3.7%; Score 320.5; DB 1; Length 1381;
Best Local Similarity 19.5%; Pred. No. 1,1e-05;
Matches 229; Conservative 183; Mismatches 456; Indels 309; Gaps 48;

OY 3 STPSKLLP-IDKSHLQLOPOSSASIFNSPTKPLNFPRNRSKPS-----LDPNSSSD 54
Db 378 SAFOAIATSRASKSPSLQDPHPOVASPAVNTQTPYQVLPQNSNGSLDLALALNPSFSSP 437
OY 55 TYTSEODEKKEKKEKKDPAFTQSPDRNFPLDLSIDIQOTIOHQOQPOPOOOLSTQDNL 114
Db 438 SPT-----KAGVYVQNTNTNPSFYDNN-NGQATL--QQOQPOQPPRLTHSSGL 483
OY 115 IDSEFSQTPMSTL-DITKONPTYDKAKNENHAPTYINTSPKSTIMKATPK--ASPKV 170
Db 484 -KKF--TP-TSNFGQSILKEEPEDQLRESSTF-----SAQPPVPKHAASSPYKR 531
OY 171 AFYVTNPRIHHYPRDNRVEEDQSOQKEDSVPEPLIQHOMKDRPSQFN-----YSDEPTNA 224
Db 532 TASTTLPVPPNPFVSFPMGAGATSAATGAAGAAGAAALGASAFSSRNAFAFKQODLFA 591
OY 225 SVPTPLPHTTKPTFFAOLINKNNEVNSEPALTDPK-----LKRENSNLSDEKVALY 278
Db 592 DGEASAOISNMTTEAMLSNNOVNSLSKQASTTNDKKSRATQELKRYETMKNSIQIKLN-N 650
OY 279 LSPITNNNSKVNSDMSHLQMLDASKNKTNENIHNSLFALKAPKNDIENPLNSTL---- 334
Db 651 LRSITHDQVKTQTEOLEAQVQV-----NKENETLAQQLAVSEAVYHAESKLINELTTDQ 705
OY 335 -----MDI--SLRSSGSSGSSLOSLRNDNRYLVESVQSPKKV-N-PGLSLND----- 378

```

Db	706	ESQTKNAALKEQITNUNLSMTASLQSLQENKQOQVKQERSNVADVNSQGLNQTNYTANLQK	765
Qy	379	GKFGSDEVSLPRDLRSRKLKLTTEHDAPEHNENNFTDAKSTN-TNKGLLYSSDHH	437
Db	766	ELDGGEEKISVYLTKQKELANDYOQTVEEHQAOLQAKYODLSNNDPTLTLTREKOLEERNQ	825
Qy	438	LDSPFRSNHFEOSLNLNLNSASQOSQSLNALNEKORQTOEQOQAAPPEEFTSFQDNK	497
Db	826	IEEQDNLVHGVSKLQEFMDLSQRKASFKADELK-----ERNETVANNR	873
Qy	498	VKQEPKSNL-----EPVYVTKKEPVSAETAKPKREFSSRLIKNEDEIAEPADHP	551
Db	874	ELSEKQMLANGQLPEDAKDIATK---SASNTPTTTEKATSR-----GNVHE	917
Qy	552	KKENF-ANSVHEEDLALKKALNDEESDTQNSTKMSITREHID-----SDMKLEDSN	603
Db	918	DTVSKFVTTVENSILNVRKADDEEKTERESDV-----FDRDVPTLQSQSSEANVN	972
Qy	604	DGDREDNDISREFKSDILNDVQSTSDIIDKKYQNSSEITTTKTLAPPSDNNDKENSKS	663
Db	973	NGTQSGNETAN---PULTELTSLDRFGDGLNEXGIPRQSLGTSSV-----	1011
Qy	664	LEDPAANN-ESLQOOLQEVPHTKEDDSLANSN-----IAPPELLLPVYEANDYS	713
Db	1014	-----ANNAPOQVRQDVLEPELTERDITNTNANDMTGLNSHLPGEWEATPASTDIVLS	1065
Qy	714	FNDVTKTFDVAVSFEESLSRHEHEDSKPINFISIMWKEQKKHQIKHVKPTKOIIASVQO	773
Db	1070	-NETTEVEIEDGSTTKRANSNE---DGEVSYSI-----QESPKISAPKAKTI----	1112
Qy	774	YKNQESNVTSDKKIKTNALQFKKFEVNVMSRRVVSPPMDLNVSGFLPELSEDSQERD	833
Db	1113	---NEE-----FPPLQELIHIDESSSSSDDDEFD	1133
Qy	834	LNFAVNSMNTNRPSRFTPLSTKRVLSINDDPVRVBEPPKSAETIRNMRILSANKAKPN	893
Db	1140	TR-----ELPSATVTKLQPPV-----AQTSSLEHT-----E	1166
Qy	894	QAPLP-PQRPQSSSTRNSNKRVRSPREVPTEIRRTSSALAPCDMYNDF-----D	943
Db	1169	QVIYKYPAGTSPSHNEGNSKK-----ASTNSILPVKDEDFEAGLEQAAVE	1216
Qy	944	DFGAGSKPTIK-----AGMKTLPSMKDDVYKRIILAKKGYTQDEYIMAKLVQDKRK	996
Db	1217	DNGADSEEEFNANAGSMQOFETIDHKDLDELQ-IMAFYGLTSS-----SS	1263
Qy	997	NSIATDEDRKEELQOQTASIHNAITIDSSIYG-----RPSDISTDMLPYLSDCLKPPTA	1050
Db	1264	NPTLPKQVQOQSNSDPAQVSNDEMDEIFGPFQNSKAETPKYATPSIQQPIPLKNDPIV	1322
Qy	1051	-----LLSADRLL---FMEQEVH	1064
Db	1324	DASLSKGPVNRGVATTPKSLAVEELSGMGFTDEEFAH	1360
RESULT 5			
MAPB_MOUSE			
AC	P14873	STANDARD	PRT: 2464 AA.
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))		
DE	[Contains: MAP1 light chain LC1].		
GN	MAP1B OR MAP1B OR MAP15.		
OS	mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND DOMAIN.		
RC	STRAIN=Swiss Webster; TISSUE=Brain;		


```

Db 1670 ASVLHTTNGPTVVDYSPCDIDSSLSHKTPPEEPSYTO-DNDLSLISVSO-VEASPS 1172
Oy 1142 SXSANTRGDDDLASAPETPRTPRKESIS-----SK 1174
Db 1728 TSSAHT-----PSQIASPLQEDLSDVVPREMSLVLASEKVSLEGKELSP 1776
Oy 1175 PAKLSASPRKSPKIKGSP-----VAVIKNGS---IAGIEPIPKATHPK---KKSPOGN 1223
Db 1777 KSDISPLTPRESS-PLYSBGFSDSTSAKETAAHQASSPPIDATAEPYCFRSMLEFD 1835
Oy 1224 EISNHNVRGGISPS--GSEHOHNPSMVSPQYTDATSTVDEKDKVQHKREKQOK 1282
Db 1836 TMOHNLALNRDLTSSVEKDSGKTDGDFNYAVQKPEMAAGSPDE-EDYDYESQEKTYIR 1894
Oy 1283 NHHNNHHNNHKKOTDIPGVVD-----EIPDVG-----LQERGLKLF 1319
Db 1895 HDVVRYYEKEKERT-IKSPCDSCSYETIEKTKTPEDGVCCEIETKTRPREBEGCY 1953
Oy 1320 RVLGITINILPDIN-----THKGRFL-TLDNGVHCVTTPREYMDHNAVIGKEFELT- 1371
Db 1954 EI-SEKTTTPREVSQYTYEKTERSRRLDDISNGYDDTDEGHTLDCSYSETTEKITS 2012
Oy 1372 --VADSLFETLTKASYEKPRGTVEVEKRYKYSNRRLSRLFGSKD-IITTKFVPEV 1428
Db 2013 FESESESYETSTKTRSDPTSAICYETMEKIKTKPOASTYETISDRCYTEKKSPSEA 2072
Oy 1429 KDTWANKFAPDGSFARCYIDLOFEDOTIGKASQF---DLNCFNEMETMSNGNQPKR- 1483
Db 2073 RQ-----DVLCLVSCGFKHPKTELSPSFINPMPLEFACEEPTSEKELTQS 2122
Oy 1484 -GRPKYKADLEKMLVPPSDR-----ELTPSIRKAYESINLNNQNNYFEGYH 1535
Db 2123 GQAP-----PPSGKQOQROGDETPPTSVSAPSQTDSD-----TP 2159
Oy 1536 QEGDCDPIKRRFKFLMGSLSLANSEISHKTRAK-INLSKVVDLIYVKENI---DRSNH 1591
Db 2160 PTEBEP-----SITADANIDSEDESERTIPTKTYTKHMDPPRPMODRSPS 2207
Oy 1592 RNFSVLLLD 1601
Db 2208 RHRDVSMD 2217

```

RESULT 6

```

67_YEAST STANDARD: PRT: 1658 AA.
YMG7_YEAST Q03661; Q04988;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
Hypothetical 187.1 kDa protein in GUA1-ERG8 Intergenic region.
YMR219W OR YMR261.13 OR YMR959.01.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
SEQUENCE OF 1-711 FROM N.A.
STRAIN=S288C / AB972;
Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
Walsh S.V.;
Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 608-1648 FROM N.A.
STRAIN=S288C / AB972;
Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

```

Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc or send an email to license@sib-sib.ch).
Cc -----
DR EMBL: 249809; CAAB9934.1; -
DR EMBL: 249939; CAA90180.1; -
DR SGD: S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match
Best Local Similarity 3.6%; Score 313; DB 1; Length 1658;
Matches 233; Conservative 198; Mismatches 498; Indels 424; Gaps 57;

Oy 42 NSKPSLDPNSSDPTVSE--ODEKGEKKKDTAFQTSFDRN----- 82
Db 440 NTSESRDGFADATYKKNVEQDEDEPEKDDIITSSLDKFNHGNKNSSEYSENVLENE 499
Oy 83 -----DL-----DNSIDI--QOTIOHQOQO---PQOQ 104
Db 500 TDPAIVERNQINDYEGVDYTGKSVESDLHESHPDLYLDAARMLQFOQRNSNCPQKE 559
Oy 105 QOLSQF-----DNL-----IDFSQTPMTSTLDTKONPYDKYENHAPTYITSPN 154
Db 560 EOVSSELYGHSNOSNLGSRSLDESEBQIPK--DETGNNNNLKTDRGLDSSVEIEVE 616
Oy 155 KSIWK---ATPK-ASPKVAFVTNPRIHNYPDNRVEEEDSQQKEDSEVPPLOHQK 210
Db 617 KYSEKKLSTKELVPLSTDTTINNSSLGNEDSIYSLDDADAISENLTDPYLMIKTT 676
Oy 211 DP-----SOFNTSD--EDTNASVPPPLHTTKPTFAOLLKKN-----VNSEPE 254
Db 677 PRYEVVISESVYSSTSYEDNTVAMP--POVEYTSFPMNDFNSLNDYEEKHDLKSTLA 734
Oy 255 ALDMLKLRK--NPSNLSLDEKVNLYLSPTN-----NSKNVSDMDSHLQNLQDASKNT 308
Db 735 ALAPAFTKDAEFVEAGVTKSCLTSGHTNIFHSKTEKQVSDLESTENFEENEG 794
Oy 309 NENIHN-----LSFALPKPKNDIENPLSLNADISLRS 342
Db 795 DENKNSKNFPGVANSTOKSTEDNTDEKVFSAINTVNGDSSCDIETASNVENLRY 854
Oy 343 SSSOSSLSLNDNRNRYLESVSPKKNVPGLSLNDGKIGFSEDEV----- 388
Db 855 CEKDNNEAEMSSGDCBCKVKNDDGSKTOISFYSDSPDNDESDNTEFSSTKYVNSDLE 914
Oy 389 -ESLPRDLR---DKLETTKEHA-----PEHNENPIDKKSTNTKGGOL 431
Db 915 DDESLKELTKAEVVDKLDDESEDSYEDDYADPREGDGSENNIV--KTKKDTLGLV 972
Oy 432 VSSDHLDSFDRSYNHTQSIINLNSASQOISLNALEKOROTO--EQEOTQAEPERE 489
Db 973 EPENEKVNKV-----HEETLFE-ANVSS--SYNVQKKDMHTDVIQDEQAENAEGR 1022
Oy 490 TFSQNIKYKOPKSNLEVKYTKIKPEVSATIEAPKRE-----FSSRLRKFN 539
Db 1023 KYIQNTDTEAHISITETI-----DENAIGNNMEIPERSVEKETHNEVLFERRATTIEN 1077
Oy 540 EDEIAEPADIIHKKEANEASHVEDTDLKALNDEESDTT-----ONSTRMSRPHD 594
Db 1078 TKAL-----ENNTNMH--DOVSQACSDSDRQDOSTAEKNVGSAKHNLDIYVS 1123
Oy 595 SD-----WKLEDSNDGREDND 612
Db 1124 SSEIESVEPLKPEDSRNIFSSPIRVIGAVYKGVVQVDAESFYKKLIDVMDSEDDVND 1183
Oy 613 ISREFKSDILN-----DVSQTSIDIIGDYGNSSS--ITTKTLAAPPSSDNKKNKNSL 664
Db 1184 IGDYNO-DIFNKSNTSDASVNMKSVSKERDSDEDEAVILGCVTEAHNDGN--NSRVI 1240
Oy 665 E-DPANNSLQOQULEVPTTKEDDSLANSNAPPEELTLPVEN-----NDYSS- 713
Db 1241 NIDPTTGAYEEDSEVPRQOYKD-----KENLHSEE--PLVEGLQSEQHEKKDHSEN 1292

```

```

OY 714 -----FNDVTKTFDAYSFEESLSREHETDSKPIINFISIMHKOEKKOHQHXVPTQ 766
DB 1293 EEEFPTIYGDITSA-NIHSNAPDDIKR-----OQLLNLSLDLEWYSOR 1334
OY 767 IIA5YQYKNEQESRVTSDKVIIPNAIQPKKEVNVMSRRVSPD-MDDLNVSOPLPET 825
DB 1335 LLEDSRKGKNEOE---SDEVMTSRERDLTFPEKSVNEKACAGAEEDTFSELDISIQHPH 1390
OY 826 SEDSGFKDLNFANYSNNTNRPSFTPLSTKNVLSINDPNVVE-----PPEPKSYA 877
DB 1391 EED-----LDLSNNOGRS-----IEELNSEPEAELEYELEIEGTFETAASS 1431
OY 878 EIRNARLSANKAA---PQAPRLPQROPSPSRNSNKRVSFRVPT---FEIRRTSS 930
DB 1432 KNNDEOROGNIPSTDLDPSPSKEEYTDSTYPSSENTIAEKSAPTSEVYEISDP 1491
OY 931 ALAPCDMYNDIPDFGAGSKPTIKAGMKTLPSPMDKDYKRLNNAKGYTODEYINAKLV 990
DB 1492 NEVPMELNDEI-----PATTLKHXDKTNTVSVD-----DRSEHLSHDY 1531
OY 991 DQPKKNSVITPDEPRYEELQOTASIHMATIDSSITGRPDSISTDMLPYLSDELAKRPTA 1050
DB 1532 DNEPHDINSINIKVNEGESEEHQAVDI-----PVKV-----EVKE----- 1565
OY 1051 ILSADRLFMEOEVHPLRSNVLVHPGAGATNSMMLPEPDPFELINSRANVSNSDNVAI 1110
DB 1566 -----EQEEMR---SKSVLEE-----QKPSMELINDKSSPENND----- 1598
OY 1111 SGNASTISFNQDMNFDQATIGQKIQEQPASKANTVRGBDDGLASABETPR 1163
DB 1599 -----ETNREKDKTKAKKSKRRKRVNSRRKRRTTEGSSAASNTKR 1639

```

RESULT 7

```

MLP1_YEAST STANDARD: PRT: 1875 AA.
ID MLP1_YEAST Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RT Mol. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RT yeast 9:1349-1354(1993).
CC -I- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -I- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -I- CAUTION: REF. 2 MISQUOTES THE GENE NAME AS "MLP1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L01992; AAA34783.1; -
DR EMBL: X73541; CAA51948.1; -
DR EMBL: Z28320; CAA82174.1; -
DR PIR: S38173; S38173.
DR SGD: S0001803; MLP1.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487
FT DOMAIN 531 1678
FT DOMAIN 1834 1866
FT CONFLICT 301 301
FT FT
SQ SEQUENCE 1875 AA; 216455 MW; 683AD34C906867 CRC64;

Query Match
Best local similarity 3.6%; Score 310; DB 1; Length 1875;
Matches 294; Conservative 254; Mismatches 579; Indels 444; Gaps 63;

OY 16 HLOLOPOSSAST-----FNSPTKPLNPFRTNSKPSLDNNS-SDTYTSEQDEK 65
DB 402 HLOLOLETFVLEHHPVPIINSFKERTDMLNNAALLLEHTSMENAKVLELNKNO 461
OY 66 KEEKKDTAFQTSFDRNFDLNSID---IQOTIOHQOQOPOQ-----QOLSOTDNLLD 116
DB 462 KLVECEENDLQTLTKRHLDCROLYLLTNSVNSDKGRLKREIQFIQIMGEDSTIT 521
OY 117 EFSFQTPMISTL-----DLTKPNPYDKVKNENHAPYINTSPKSIKKATPKASPK 169
DB 522 ESDSKVETRLVEFKNIILOLQERNAELKVBNLA-----DKLESKEKSKOSLOK 573
OY 170 VAPTVNPEIHHYVDNRVEEDOSQOKEDSEVPEPLIOHOMKDSOFNVEDOTNAYSPT 229
DB 574 ISEETVNEAKKALITLKSEMDLESREE-----LOKELEF-LKTSVPNEDASYAVTI 626
OY 230 PRLHTTK-----PTFAQLLNKNE-----VNSEPALD-----MKLKRENSN 268
DB 627 KLTETKRDLESQVODLQTRISQITRESTENMSLNKEIOLDVSKDISIKLGKESR 686
OY 269 LSLDEKVNLY-----LSPITNNNSKNVSDMSHLQNLQDASKKTENTH----- 313
DB 687 ILAEERFKLSNLTDLTKAENDOLRRFD---LVQNTILKQDKTHTLEEVYCSKLS 743
OY 314 -----NLSEFALKAP---KNDEIENPLNSLTNADISLRSSGSSQSSLOSIRNDRVLESV 363
DB 744 IVTELLNLKEQKLRVHLEKNLKQELNKLSPKDSLRINWVLOTLQAKRED--LEEFT 801
OY 364 PGS-PRKVNPGSLNDGKGFSDVEVESLLPRDLSRDKLETT--KEH---DAPEHNENF- 417
DB 802 RKSCQKKI-----DELEDA-----LSLKKETSKQDHHIKQLEEDNNSNIE 842
OY 418 -----IDAKSTYTNKGQLLVSSDDHLDSFDRSYNHT----- 448
DB 843 WYONKTEALKDYESVITYVSQKOTDEKLOQYVKSLSKEKIEEDKIRLHTYVWDETIND 902
OY 449 -----EOSILNLNLSASQSOISLNALEKOROTOBOEQOAAE-----PEEE 489
DB 903 DSKRLKELEKSKIMLTDVYSOIKKEYKDYETTSQSLQOTNSKLDSFKDPTNQTINKLTDER 962
OY 490 TFSFDMNKVQOEPRSN---LEFVKVYIKKEPV-----SATEIKAPREFESSRI 534
DB 963 TSLIEDKISLKEQMFNPNLNNDLQKKGMKEKADFKKRISILQNNKNEVAVAYSEYSKL 1022
OY 535 LRIKNE-DE---IAEPADHPKKEENANSVEDTALLKAL----- 572
DB 1023 SKIQNDLDOQTIYANFAQNNYVEDELQKHADVSTISLRQLHTYKQVYTLNLSRDOL 1082
OY 573 NDEESPTQNSIKMSIRFID--SDMKLEDSND-----GDREDDNDISREF 617
DB 1083 NALKENKSVSSKESLEQLDLSNSRIEQLSSQNKLLYQIOIYTAADKEVNNSTNGPG 1142
OY 618 KSDILNDVSGTSLIGDKYNSSSSEITTKTL-----APPSDNN--DKENSK 662

```



```

Db 801 TGATTAAGAAAGIAIGPAKLEERSLMSBEDLT-----KDEELKAE-VDV--- 850
QY 281 PTNNNNKVVSMDSHLOLADASKKNTMENIHNLSFALKAPKNDLENLNTNADISL 340
Db 851 -----TKDIKPOLLEIDEBEKLETEPEV--AVIÖKREVTGKGPRESDEITTT 899
QY 341 RSSGSSOSSLQSLRNDNRVLESPGSKRVN-----GLSLMDGKGFSD--VVESLTP 393
Db 900 BEGECEQ-----PRELEPEKGV---DIEKFEDGAGFEESSEI 939
QY 394 RDLNRDLKLETTKHDAPENNNENFIDAKSTNTKGLLVSDHLDSPRSYNHTQSLI 453
Db 940 GUY-EKKA-E-TEAEPEDEGEHVCVSAKSHPTEDESAKAADAIR---EKRESVA 994
QY 454 NLNLSASQSOIQLMLEKROQEOQOAAEPEETSFDNKKVQEPK--SNLEFVAV 511
Db 995 SODDRAEEDMD--EALKEKADOSE--EADDEDKADAREEYEPKMAEYVMA 1047
QY 512 TTK-----EPVATETIKAPKREFSSRLIKNEDEIAEPADIHPRKENE 536
Db 1048 VYDKAAEAGAEQYGLTTPKQLOAQSPGRPASSI-----HDE-----TLPGSESE 1097
QY 557 ANSHVEDTALLKALNDEESDTONS-----TMSIRFHIDSPWKE 600
Db 1098 ATASDEEN-----REDQPEFTATSGYTOSTIEISSEPTPMDEMSTPRDVMSETNN 1149
QY 601 DANDGREDNDISREKSDIINDVSOTSDI-----GPKYGSSEIT 645
Db 1150 EETESSQGFVNTTKESSLYSOEYSKPADVPLNGFSECKTDAIDGADYNASAS--- 1205
QY 646 KTLAPRSDNDNKENSLEDPANNE--SLQOOLEVPHTEKEDSLANS----- 693
Db 1206 -TISPPSSMEEDKFSALRDAVCSEVKASTLIDI-----KDSIASVSSKVSPPSKSPL 1259
QY 694 NTAPEEL-TLPEVEANDYSF--NDVYTFDA-----YSEFESLSREH--ETDSKPIN 743
Db 1260 SPSPSPLEKTPLEGRSVNFSLTPNEIKVSAEAVAVSEVYQVEVEHCASPEDKTE 1319
QY 744 FLSIMHKOKKKHOKHOKPTKOILAS-----YQYKNDQESRVSDVKIPNAT---- 793
Db 1320 VVS-----PSQSVTGSAGHTPYQSPTEKSSHLTEVLEKPAVVSF 1363
QY 794 QFKKEKVVMSRRVYSPMDMLNSQFLPELSEDSGFKDLNPNANSNNTNRRSTPLS 853
Db 1364 EFSDAKDN--ERASVSP-MDE-----PVDSSEPI- 1391
QY 854 TKNVLSINDNPNVPEPEPKSYAETIRNAKRLSANKAARNQAPPLPPQRPSSSTRSNSK 913
Db 1392 -EKVLSPLRSPPLIGESAYESF-----LSADKASGRGAESPFEK--SGKQSPD 1440
QY 914 RVSRRFVPTFEIIRTSALAPCDMYND-----IFDDGASCKPTIKAEKMTLP 962
Db 1441 QVS-----PVSEMTSTS-----LYODKOBGKSTFPAIKEDYGO-EKKTDDVEANSSOP 1488
QY 963 SMD-----KDVYKRLNNAKGVYQDEYINAKLVDQPKKNSIV 1000
Db 1489 ALALDERKLGDVSPQIDVSOEGSEKEDPK--MSISEGYTSDK--SATPVDEGV----- 1538
QY 1001 TPDEPRYEELQOTASIHNAITDSSIGRPDSITDMLPYLSDELKRP-----PTAL 1051
Db 1539 --AEDYTSHMEGVASVSTASVTSF--PEPTDDVSPSLHAIEGVSHSTEVDSLSVSV 1594
QY 1052 LSADRLFMEOEVLPLASNVLVHFGAATNNSMLPEPDELINSPARN--VSNNSDNA 1109
Db 1595 VQPTPTTFOETEMSPKEE-----CPRPMSISPPDSFPTAISRTPVQDRSQSS 1644
QY 1110 ISGNASTISFNQ-----LDMNFD-----DOATIGOKI---QEPASKSANTVRGDDGL 1155
Db 1645 MS-----IEFGQSPSQSLAMDSPSRQSPHPTVAGAGVLHTTEGPRPEVDVSPDMQDSSL 1699
QY 1156 A-----SAPETPRPTKESISSKPAKUS--SAPRKSPIKIGSPVR----- 1195
Db 1700 SHKIPMEBPSYQDNDLSELISVQVEASPTSSAHTPSQIASPLQEDTLSDVAPPRDM 1759

```

```

QY 1196 -----VIKNGSIAGIEPIPKATPHKPKKSFQGNISNKKVRD--GGISPPSGSEHOQN 1247
Db 1760 SLVASTLSEKVOSELEBKLSPKSDISPLTPRESSPLSPFFSOSTSAVKKKTATCHSSSS 1819
QY 1248 PSM-----VAVPSQYDTATSTVPDENKDVOKPREKO 1279
Db 1820 PPIDASAEPEYGRASVLPFTMQHLNLRDLSTPLGLEKSGSKTPQDPFYAYOKPEET 1879
QY 1280 KÖKHNHNHNHNHNKÖT--DIPGVVDEI-----PDVGLQ-----ERKGLF 1318
Db 1880 RSPDEEYDEYSEYKTRISDVGGYEEKLEFRTKSPSDSGSYETIGTKTPEDGDYS 1939
QY 1319 FRV-----LGIKINLPDIN-----THKGRFTL-----TLONGV 1348
Db 1940 YEIIETKTRPEEGSYDISSEKTTSPPEVSGSYEKTERTSKRLDIDISNGYDSEDEGH 1999
QY 1349 CVTTPPEYMDHNAVIGKEFELTVADSLFELTLKASYEKPRGTLYEVTEKKVYKSRNL 1408
Db 2000 TLGDPSYSY-----TEKITSPESEEGSYETSKTTRPTDTSYCYETAKEKITRTPQAS 2055
QY 1409 SRLFGSKDI-TITTKRYPTVYKDTWANKFAPDGSFANCYIDLQOFEQITGKASQF----- 1463
Db 2056 TYSYETSDLCYTAEKKSPEARO-----DVDCLVSSCEYKHKPTLSPSFIPNP 2105
QY 1464 DLNCFNENE-----TMSNGNOPMKRGKPYKIAOLEVKMLYVPRSDPREILPTIRS 1514
Db 2106 PLMFASSEPTSESEKPLTOSGAPPPRGK--QOGRCD-----ETPPSVSE 2152
QY 1515 AYESINELNNEONNVEGYLHOEGDCPIFKKFFKMLKSTSLANSEISHKTRAK-INLS 1573
Db 2153 SAVSQTDSD-----VPPETEBCP-----STADANIDSEDESEITPTD 2190
QY 1574 KYVDLIYVKENI-----DRSNHNFSDVLLD 1601
Db 2191 KVTYKHMDDPPAPVQDRSPSPRHDPVSMVD 2221

RESULT 9
USOL_YEAST
ID USOL_YEAST STANDARD: PRT: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Intracellular protein transport protein USOL.
OS USOL OR INTL OR YDL058W.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RN SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE

```

CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54378; CAA38253.1; -
CC EMBL: L03188; AAB00143.1; -
CC EMBL: U53668; AAB6659.1; -
CC PIR: A38455; A38455.
CC SGD: S0002216; USOL.
CC InterPro: IPR002017; Spectrin.
CC -----
CC Transp: Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
CC FT DOMAIN 1 724 GLOBULAR HEAD.
CC FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
CC FT DOMAIN 991 1790 CHARGED (HYPER-HYDROPHILIC).
CC FT DOMAIN 1172 1786 DISPENSABLE FOR THE PROTEIN FUNCTION.
CC FT DOMAIN 847 847 ASP/GLU-RICH (ACIDIC).
CC FT CONFLICT 924 924 G -> E (IN REF. 2).
CC FT CONFLICT 1253 1253 V -> I (IN REF. 2).
CC FT CONFLICT 1319 1319 I -> V (IN REF. 2).
CC FT CONFLICT 1461 1461 N -> S (IN REF. 2).
CC FT CONFLICT 1581 1581 G -> S (IN REF. 2).
CC FT CONFLICT 1600 1600 I -> V (IN REF. 2).
CC FT CONFLICT 1661 1661 R -> S (IN REF. 2).
CC FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).
CC FT CONFLICT 1790 AA; 206424 MW; 6CE2B216EPD4818 CRC64;
CC SO SEQUENCE

Query Match 3.4%; Score 292.5; DB 1; Length 1790;
Best Local Similarity 17.0%; Pred. No. 0.0019;
Matches 244; Conservative 254; Mismatches 484; Indels 455; Gaps 57;

50 NSSPTYSSEDOEGKEKKDTAFOTSPDRN-FDLNSIDIOQT----- 94
462 NGSKSKESKESDCKDTGCKDTEGSGFKANLEFVLYADMLNPPKLEFTTDIDMF 521
95 -QHQQQPPQQOQOQSQ--TDNNLIDFSFQTPMTSLDLTKONPTVD-KVENNHAPTY 148
522 FFOQDHKKSEELREITRNVTGNOLEDEPLKAIGTISELTSLADIRIPISVLPFL 581
149 I-----NTSPKNSIMKATPKKSPKVAFTVNPETLHHYPRDNVEEDOS---- 193
582 IYWLFGDKATNDFLSDSKSVIKSL-----LSFSY-----QIODEDVTIKCL 622
194 -----OQKEDSVPEPLIOHOM-----KDP-----SQFN-----YSDDETASVVP 228
623 VTMLLGAVAFVSSKSPKSPKRYEFETIKTKGKNYASRIKOFKDKDSIFSKVMDNSTL 682
229 TPEPLHTT-----KPTFAQLLNKN----- 246
683 TPELDETGLPKYFSTYFIQLENNIYRIATLSHDPEEPISTKISFEVEKLOQCQTKL 742
247 -----NEVNSEPEALTDMLKRENFNSLSLDEKVNLYLSPTNNNNSKNVSDMSHLQ 298
743 KGEITSLOTETESTHENLEKLIATLN-EHKELDEKYOI-LNSSHSLKENSISLETTELK 800
299 NLQDA-----SKNKTNE-----NHNLSFALKAPKNDIENPLNSLTNADI 338
801 NVROSLDMETQLRDVLETKDENQALALEYKSTIHKQEDSITLTKGLETTLISOKKKAD 860
339 SLRSGSSQSSSLQSLRNNRYLESVPGSPK-----KVNPGSLNDGIGFSDVEVESL-- 391
861 GINKMGKDLFAL-----SREMQAVEENCKNLQKREKDSNVNHQKETSKLADIAKITE 914

QY 392 ---LPRDLSRDKLETTKEHDAPEHNNENFIDAKSTNTNGQLLVSSDDHLDSFDRSYN-- 446
DB 915 IKAINEINLEEMKIQCNNSKKEKHEHLSKELVEYKSRQSDHNLVAKLTLEKSLANNYKRM 974
QY 447 --HTGOSTINLNSASOSISINALEKO--ROTOEOOTQAAPEPEEFSFSNIVKQEP 502
DB 975 QAENESLKAVEESKNESSTIOLSNLQNKIDMSQOEKENQOI-----ERGISXENIOLKKT 1030
QY 503 KNLLEFVKTIKKEPVSATKAPREFSSRLIRKNEDEIAPPD----- 548
DB 1031 ISDL-----QTEELISKSD---SKDEYSQISLKEKLETATTANDENVANKISELTCTR 1084
QY 549 -----IHPKKEANSVHEDTDLKALNDE-----ESDTQNSTKM-SIR 590
DB 1085 EELEAEALAYKNLKNLELTLETSKALKVEKNEBHLEKEETIQLEKATEETKQOLNSLR 1144
QY 591 PHIDSDMKLESDNDC-----DREDNDISREFKSDIANDVSGTSDIGDK 635
DB 1145 ANLES---LEKHEHDLAOLKRYEEOIANKEROYNEELISQI--NDEITSTQOENESIKKK 1199
QY 636 YGNSSEITTKTLAPRSDNNDKENSKSLDEPANNESLOQOLEVPHTKEDDSITLANSNT 695
DB 1200 --NDELEGEVKA--KSTSEOSNLKSEIDALNLQI-KELKKNETNEASLESISYV 1253
QY 696 APPEELTLPVEANDYSEFNDVTKTFDAYSSPEESLSREHETDSKPNFISIMHKOEQK 755
DB 1254 ---ESBTYIKRLEQECNKE-----KEVSELEDKIKASEDKNSKYL-----ELQK 1296
QY 756 KQIHKVPKQIILASYQCKNQEBSRYVSDKYNALQFKFKE--VNVMSRRVYSPDM 813
DB 1297 ES-----EKKEELDAKTELKIQLEKTNLSAKKESSELSRLKTTSE 1342
QY 814 DDLNVSQFLPELSEDSGEFQDLNFANYSNNTNRPSTPLSTKNVLSINIDNPVVEPEEP 873
DB 1343 EKKNAEEOLEKLNELQIKNOAF-----EKERKLNGSGSTITIQEYSEKINTLE---- 1391
QY 874 KSYAETIRNARRLSANKAPNOAPPLPPOKOPBSTNSNKKRVSRVRVPEETIRRTSSALA 933
DB 1392 DELIRLQENNELKAK-----EIDNTRSELE 1416
QY 934 PCDMYNDIEDDGGASPKPTIKAGMKTLPSSMD-----KDVKR-----ILNKKKGVTD- 982
DB 1417 KVSLSND-----ELLEKONTIKSLQDELISYKDKITTRNDEKILSTIERONKRL 1465
QY 983 EYINAKLVQDKPKKNSIYTPDERYEELQOTASIHATIDSSIYGRPSISTDMLPYLSD 1042
DB 1466 ESKLQDLRAQESKAKV-----BEGKLKLEBESSKEKAELEK-----SKEMMKLES 1512
QY 1043 ELKRPPTAL-----LSADRLFMEQEVHPLR-SNVLVHPGAGATNSMLPE 1088
DB 1513 TIESNETELKSSMETIRKSDKLEQSKSAEEDIKNLQHEKSDLI--SRINES--E 1564
QY 1089 PDEFELNSPAR-----NVSNNSDNVALISGASTISFNOLD----- 1123
DB 1565 KIDBELKSKRLIEAKSGSELETYKQELNNAQEKIRINAEENTVLASKLEDIERELKDKQA 1624
QY 1124 ---MNFDOATIGQKIOE-----OPASKANTVRG----- 1150
DB 1625 EIKSQOEKELTSLRKLQELQDSTQOKAKQSEERRAEVKKFQVEKSQDEKAMLEET 1684
QY 1151 -----DDGLASAPETPRPTTK-----ESISSPAKLSSASPRKSPI 1188
DB 1685 KYNDLVNKEQAMKRDREDYVTKTTDSORQIETKLAKELDLKAKENSKLKEANEDRSEI 1741

RESULT 10
TANA_XENLA
ID TANA_XENLA STANDARD; PRT: 1744 AA.
AC 001550;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tanabla.

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tadpole head;
RX MEDLINE=92398961; PubMed=1524825;
RA Hemmati-Brivanlou A, Mann R.W., Harland R.M.:
RT "A protein expressed in the growth cones of embryonic vertebrate
RL neurons defines a new class of intermediate filament protein.";
Neuron 9:417-428(1992).
CC -I- TISSUE SPECIFICITY: GROWTH CONES OF EMBRYONIC VERTEBRATE NEURONS.
CC -I- DEVELOPMENTAL STAGE: IS EXPRESSED IN THE NEURULA AND PERSIST
CC DURING EMBRYOGENESIS IN THE BRAIN, CRANIAL NERVES, AND SPINAL
CC CORD.
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M99387; AAA4966.1; -
DR PIR: JH0720; JH0720.
DR InterPro: IPR01664; IF.
DR Pfam: PF00038; filament. 1.
DR PROSITE: PS00226; IF. 1.
KM Intermediate filament: Coiled coil; Neurone.
FT DOMAIN 1 12 HEAD.
FT DOMAIN 13 314 ROD.
FT DOMAIN 315 1744 TAIL.
FT DOMAIN 8 48 COIL 1A.
FT DOMAIN 49 60 LINKER 1.
FT DOMAIN 61 156 COIL 1B.
FT DOMAIN 157 179 LINKER 12.
FT DOMAIN 180 193 COIL 2A.
FT DOMAIN 194 199 LINKER 2.
FT DOMAIN 200 314 COIL 2B.
SQ SEQUENCE 1744 AA; 199561 MW; 6502EAC9F6C4E93 CRC64;
Query March 3.4%; Score 291; DB 1; Length 1744;
Best Local Similarity 18.6%; Pred. NO. 0.00021;
Matches 303; Conservative 258; Mismatches 614; Indels 458; Gaps 77;
QY 68 EKKDPAFOTSFDRNEDLNSI-----DIQOTIOHQOQPOOQOOLSTQDNLI 115
b 314 EAESRIRITDYGSTYFNDMSLEHNNVRRROSEDRKTYSKHRSYSKKQIG--DKNKL 371
QY 116 DEFS---FOTPMTS-----TLDLTQONPTVDKVENHAFYINTSPNKSIMKATPKASP 167
Db 372 QRPSTLNNSTYKSSAVPVRTSPVTKEFQKVSSVLSOGILKTKAPQVKEQVSTYK-- 428
QY 168 KKVAFVTVNPEIHNY-----PDMRVEEDQ-----SQQKEDSVEP 202
Db 429 -----SNLEHTFTSGAIFRAQVETKRTDQVIAIKDALGLINDLNKNGFEKEEDIQRP 481
QY 203 PLIQHQMDSQFNSDSDDTNASTVPPTPLHTTKPTFAOLLNK-----NNEVNSEPEALT 257
Db 482 GFMDF-----VSKSVSTEHKVPEDIPDLBSALSKLEEDLSVSTFPAAGSSNLEAIK 535
QY 258 DM---KIKRENFSLDEK---VNLVLSPTNNNSKNVSDMDSLQNLQDASKNKTN 310
Db 536 DVLGEPICLENLQNEIAFEKESPGTNAADPIEEVIESVSQYVHFKE-QELSNLEIE 594
QY 311 NIHNLSFLAKPKNDIENPLNLTNADISLRSSGSSQSLQLRDN-----RYLSEV 363
Db 595 NTH-----ENHVODATQAFNSCEDGDHGRAS--TLENNPEVDQVQYIRTL-- 637

QY 364 PGSPKKNVGLSLNDGKSGFDEVESSLPRDLRSKLETKENDHAPENNENFIDAKST 423
Db 638 -----SNEIKESKIPSD-----NTEBAELISRSRVLE----- 666
QY 424 NTNKGOLVSPDHLDSFDRSTNHTEQSLTNLNSASQSLNLALEKORQOEQOQOA 483
Db 667 ---NEYIPVSKDDLTEFT---SHLE-----NDESSQSFDSKLFENKSTEDQLITNL 712
QY 484 APEEETSFSDNIKVKQOEKSNLEFYKV---TIKKRPVATETKAKRREFSILIRKNE 540
Db 713 KNTQENIFQSN---QEHLENLEFDSVPDVTYK-----FMYPOE 748
QY 541 DEIAEPADIIHPKKE-----NEANSHEVDLALKLNDDESDTTONKSKIRPH-ID 594
Db 749 NNLLEENYVGGGELQVQMTDENIINQSSDOLL---LSHSHHEETKTESIAVENRNE 805
QY 595 SDW-KLEDSNDGDRDNDDIS---REFKSDILNDVSQT--SDIIGDKYGNSSSEITTKT 647
Db 806 SEHAEDKSSSEIPEVISENVSVETIHEISDVEEDTKQAFEDERVGEQINQNNQE--ST 862
QY 648 LAPPSDNDNKNKSLDEPAN-NESTLOQQLVPHTKEDDSILANSNTIA-PPELTLYV 705
Db 863 VLDGVSYSQERNQLEDEVSISEQIEKDFEI---NEQECKSDQIREAPFTEEDHVOY 919
QY 706 VANDYSSE-----NDVTKTFDAVSSPEESLSREHETDSKPINFISIMHKOEK--OK 755
Db 920 VFMQEQSFREVQGLNNIKQEVDTLQNTDED--SFQNNDEPQELSCDLOEQIKLEE 976
QY 756 KHIQIHVPYKQIIA-----SYQYKNEQESRVTSIDKVIIPNAIQFKKFEVNVMSRV 808
Db 977 ENQISENEGQNFNGNDIEEFSGQGYDTDEIQET-----IGNQVSAQLCESDINDQKL 1031
QY 809 VSPMDDLNVSQLPELSDSGFKDLNFANYSNNTRRPSFPLSKVNLSTINDPNVY 868
Db 1032 SMEDEEON-----NPETENIGLEQ---ESDQENTRSNEGTFKFS-----QEECDV 1075
QY 869 EPEPEKSYAEIRNARLSANKAAPNOAPPLPPORQPSRSRNSKVSFRVPTFEIRRT 928
Db 1076 FKPED-----MSKSEYSGQGEDLDQVYDFSLBEQANNDL-----LEKE 1115
QY 929 SSALAPCOMYNDIFDDFGAGSKPTIKAEQMKTLPSMDKQ---DYKRLINAK--GVYQ 981
Db 1116 EYILHHDQQRVSNVEITIDEKLSERI-----IDNELATGVNESAANKQVADLPT 1167
QY 982 DEVI---NAKLYD---QPKKNSITVPDEDKRE--ELQOTASINHA-----TIDSSY 1026
Db 1168 DEYAADNDYGMDDDSQYQTKEDLFVDDGNNTIEKIEIQOTSLNLQICERVNDVEDIS 1227
QY 1027 GRPDSISTMLPYLSDDELKPPALSLADRLPMEQEVNPLRSNSVUVHAGAGATSSML 1086
Db 1228 GEAKNESVEM---NDVVDLYPEAKVTGD---EQISPLQDEKLLKLEFMEDTKMDGQL 1278
QY 1087 -----PEPDF-BLINSF--ARNVSNNSDNVAISGNASTISFNOLDNPNDDQAT--TGOK 1135
Db 1279 CLEKENETEYIEVTDSPQFATLSDH-----AGRELTVQDANSANQFCENPFTKLIANH 1332
QY 1136 IOEQAPKASNTVVRGDDGLAS---APEPRPRPK-----KESISSKAPKL 1178
Db 1333 IEYETVA-----DSLESTEEQOVQETERTPEFRPESDKMENENSESESVDSQELSL 1383
QY 1179 SSASPRKSPKIKIGSPVYVIRKKNKSGIAGIEPIPAKTRPKKSFQGNISNHKVVDGISPS 1238
Db 1384 NSHKSEFE-----ISKDYQLE--QTLRVYTLPLNLEDEFEDLTE-----QPD 1424
QY 1239 SGSEHQINPMSVSPQYTDATSTVPDENKDVQNRKREKQKNNHNNHNNHNNKOKTDI 1298
Db 1425 VHEEHQNDSDGAS---TFITSVDEDEKEREVESVSKDEESN----- 1463
QY 1299 PCGVNDEITPDVGQEGKGLFFRVLGKIKNINLPDIINHKGKRFITLQNGVNCVTPPEYMD 1358
Db 1464 ---EEFEGD-----VLSVDKTSQVEVTTLSG-----LAQEPSTLQD 1496
QY 1359 DHNVAIGKEFELTVADSLFEIILTKASYEKPRCTIV-----EYTEKKVY----- 1402


```

QY 785 DKVILPNAIOFKKFEVNVNSRVVSPDMDL-NVSOELP-----EL 825
Db 1661 EIEIEN-IRLTQILHENLEMKSVTKERDLSVEETLKVERDQKEMLRITITDLEK 1719
QY 826 SEDSGFKDLNFANYSNNTRRPSFTPLSTKNVLSNIDND-----PNVPEPEPKSAEIRN 881
Db 1720 QELKIVHNLKHEQETIDKLGIVSEKT-NEISNMOKLEHNSDLKQODLKIOELK- 1777
QY 882 ARLSANKAPNOAPPLPPOQOPSSSTRSNKRVSEFRVPTETIRTSALAPCDMYNDI 941
Db 1778 ----IAHMLKQOEITIDLRGIVSEKTDLSNMOK-----DLEN-- 1813
QY 942 FDDFGSGAPTKAGGKT-----LPESMDQVKRLNNAKGVYODEYINAKLVDPKKN 997
Db 1814 ----SNAKLOEKIOELKNEHOLITLKD-----VNETOKKVSSEMOLKKQIDQS---- 1860
QY 998 SIYTPDEPDEYELQOTASIH-NATIDSSITYGRDSSIS--TDMPLYSDELKPPALLSA 1054
Db 1861 --LTLSKLEIENMLNLAQELHENLEMKSVKMERDNLRVVEETLKLERDQKESLOETKAR 1918
QY 1055 DLFMEQEVNPLRNSVLNPRGAGATNSSMLPEPPELINSFARNVSNNSDNVAISGNA 1114
Db 1919 D-LEIQOELKTAR-----MLSKHEKETVVDKLRKISE----- 1949
QY 1115 STISFNQDLMNFDQATIGOKIOEOPASKAMTVRGDDGLASAPETPTPTPKESISSK 1174
Db 1950 KTIQISDIOKDKSDKDELQKTIQOELQKELDOLLRVKEVNNHSHKINENEOLOKPEPN 2009
QY 1175 ---PAKSSASPRKSPKIGSPVRYVKNGSIAIEPIPKATHPKPSF-----QONE 1224
Db 2010 YLCKEMDMFOJLTKHNELEIRIYAKERD--ELARIKESLKMEDQFIATIREMIAND 2067
QY 1225 ISNHKVRDGGISPPSSSEHQDN-----PSMVSVPQYDATSTYDENKDVO 1272
Db 2068 RONHGVKRRKRLYLSDOQOIMESLREKCSRIKELLYKRVSEMDHYECLNRLSIDLEKETE 2127
QY 1273 -HKPREKOK-----OKNNHNNH-----HHHKKOTIDRGV 1302
Db 2128 FRIKAKKLVYLVYTKIEBOHECINKFEMDFIDEVEKOKELLITQILOQDCVDP--- 2184
QY 1303 DDEIPDVQIQRGKLEFRVLGKININLPDINTHKGRTLLDNGVHCVTTP-----YN 1356
Db 2185 SRELRLKLNQNMMDLHIEEI-LKDFSESEPPSIKTEFOQVLSNRKEMTOGLEWMLNTRFD 2243
QY 1357 MODHNAIQRKEFELYADSEF-----ILTAKASYEKRGITVETTEKKVYVSRNLSR 1410
Db 2244 IEKLNKIGOKERDRCQVNNFNNRIIAINESTEFEEERSATISKEMEDLKSLEKENEK 2303
QY 1411 LFSGKDIITTKFVPTLEVADTAMKFAPDGSFARCYIDLQOFEDQITGKASQFDLNCFNE 1470
Db 2304 LFKNVOTLKTSLASGQVAPPTQDNKP-----HTSRAOTLTTEKIRE 2347
QY 1471 WETMSGNONGPKRGPKYKLAQLEVKMLYVPRSDPREILPTSIKSAVESINELNN--EQNN 1528
Db 2348 LENSLSHAKESAMHKESKIIKQKEL-----EVTNDIIAKLOAKVHESNCKLEKTK 2398
QY 1529 YFEGYIHOGS-----DCPIFKKRFKL-----MGTSLAISEISHKTRAKINLSKV 1576
Db 2399 ETIYQVLDVVALGAKPYKEIEDLKMKLVKIDLEKMKNAKEFEKES-ATKATVEYOK-- 2455
QY 1577 DLIVYDENKIDRSHNHFSDVL----- 1600
Db 2456 EYIRLLREULRSQOQODTSVISEHTDPOPSNKPPLCCGGSGIVQMTKALILKSEHRL 2515
QY 1601 DHAFTIKFANGELI-----DFCAPNKH--EMKITW 1627
Db 2516 KEISIKLQONEOLIKOKNELLSNNQHLISNEVKTW 2549

```

```

AC P70478:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein).
GN APC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fischer 344/N; TISSUE=Brain;
RX MEDLINE=96116966; PubMed=8563176;
RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
RA Sugimura T., Nagao M.;
RT "CDNA Cloning of the rat APC gene and assignment to chromosome 18.";
RN [2]
RP Mamm. Genome 6:746-748(1995).
RC STRAIN=Sprague-Dawley, and FISCHER 344/N;
RX MEDLINE=95148647; PubMed=7846077;
RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
RA Weisburger J.H., Sugimura T., Nagao M.;
RT "Specific 5'-GGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
RT tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
RN [3]
RP Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
RC CATEININ, APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC -I- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATEININ (BY
CC SIMILARITY).
CC -I- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
CC SIMILARITY).
CC -I- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -I- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: D38629; BAA07609.1; -
DR HSSP: Q02248; 3BCT.
DR InterPro: IPR000225; Armadillo.
DR Pfam: PF00514; Armadillo_seg; 6.
DR SMART: SM00185; ARM; 5.
DR PROSITE: PS50176; ARM_REPEAT; 1.
KW Anti-oncogene; Phosphorylation; Coiled coil; Repeat.
FT DOMAIN 1 728 LEU-RICH.
FT DOMAIN 125 62 COILED COIL (POTENTIAL).
FT DOMAIN 451 493 COILED COIL (POTENTIAL).
FT REPEAT 503 545 ARM 1.
FT REPEAT 546 589 ARM 2.
FT REPEAT 590 636 ARM 3.
FT REPEAT 637 681 ARM 4.
FT REPEAT 682 723 ARM 5.
FT REPEAT 724 765 ARM 6.
FT REPEAT 766 808 ARM 7.
FT DOMAIN 739 2831 SER-RICH.
FT DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1864 1891 HIGHLY CHARGED.
FT MUTAGEN 523 523 C->R: IN AN IO-INDUCED COLON TUMOR.
SQ SEQUENCE 2842 AA; 310530 MW; 3CB32EA8A34E8F47 CRC64;

```

Query Match 3.4%; Score 289.5; DB 1; Length 2842;

Best Local Similarity 20.3%; Pred. No. 0.00043;

Matches 295; Conservative 188; Mismatches 514; Indels 457; Gaps 68;

QY 7 KLLP----IDKSHILOPOSSASIFNSPTKPLNFPRTNSKPSL-----DPNSSSDYVTS 58

D	b	1614	KLLPQSRLQAQKHVSFTPGDDVPRVYCEGTPIENFSTATSLSDLTIESPPNELAAGDV	1673
Q	y	59	EDDQKGEKEKKDT-----AFQSPFRNEDLONSIDIQOT	94
D	b	1674	RASVSGEFERDITPIEBGSTDQAQKYSIAIPDLGSAAREEDLIAECINSALPKG	1733
Q	y	95	-OHQ-----QQPQQQQQLSOTDNNLLDEFESFOTPMYSTLITLTKOMP-----T	136
D	b	1734	RSRHRFRVYKLTMDVQQAQSMSTSSTKNQJD--TKKKFTSPYKPRPONTETRTVYRKNT	1791
Q	y	137	VDRKNENHAPYINTSPNKSIMKAPKPAKPKV-----AFTVNPETIHHYF----	183
D	b	1792	DSKVVNTEETFSQDNKSKKQSLKNNPKDLNDLTPONEDRVRGFTPPSP--HHYAPTEG	1849
Q	y	184	-----DNKEEEDSOQK-----EDSVEPPLIQHQMKDPSQFNYSDE	220
D	b	1850	TPVCFGRNDSLSSLDFFDDDDVDLSREKAEILRKGEKSDS-EAKVYCH--TEBSSSQOASR	1906
Q	y	221	DTNAGSVP-----TPPLHTTKPPFPAOLKNNEVSEAEALITDMKLRKNSNLSDKE	274
D	b	1907	KAQASTKHVPVRGSKLLOEQPTFPQ--SKDVPDGAATDKLO-----NFALE--	1955
Q	y	275	VNLYLPTNNNSKNVSDMDSHLQNDQASKNTNENIHLSPALKAPNDIENPLN--	332
D	b	1956	-NTPVCFGRNSSLSSLDVD-----QENNNNE-----TGVRAAE-PANAG	1996
Q	y	333	-----LTNADILSRSSGSSOSSLRLRDNVLESVGS--PKVNPGLS	375
D	b	1997	QPGKPAQSGVAPSFHEHDTPVCF--SRNSSLSSLSIDSEDLARCISANAPKKRPRSL	2055
Q	y	376	LNDIGKGFSEVVESELLPRDLRDKLETTKEHDAPEH-----NENETDKNTNTKNGQ	429
D	b	2056	KGEG-EMQSPRKKVSVLAEDLTLD-----LKDQPRESEIGLSPDENF-DMKALIOEGANS	2109
Q	y	430	LVVSSDDHLDS--FDRSYNHTEOSILNLLNSASQSLSNALAEKOROTOEOQTOAAP	486
D	b	2110	IVSSLHOAAAAACLSRQASDSDSITSL-----KSGVSLG--PFHLTPQEE-----	2156
Q	y	487	EEETSFSDNITKIVQEPKSNLEFYKVTIKKEPVSATETIAPKREESRLL-RIKNEDELA	544
D	b	2157	KPFTSHGPRILKPRGEKSTLEAKRISENK-----GIKGGKVYKSLITGKIRSENSEIS	2211
Q	y	545	-----EPAD-----IH--PKEMEANSYHEDDTLALK-----KALNDDSEOTQ	582
D	b	2212	QMKRPLTOTNMPSTSRGRTMHIITGVNNSSSSTSPVSKGPRLPKTPAKSPBEGVATTPS	2271
Q	y	583	NSTKMSIRPHIDMDKLEDSNDGREDNDNISRFKSDILNDVQSTQSIIDOKGNSSE	642
D	b	2272	RGTKPAV-----KESLSPITRQTSIISSNKG-----	2298
Q	y	643	ITTYTLAPPSDNNKDKENSKSLEDPAANNESLQQQLLEVPHTEEDSITLANSNIAPEELT	702
D	b	2299	-----PSRSGGRSDTSPSKPTQOP-----LSRPQMSGGR--NISISGRNGISTPKNLS	2343
Q	y	703	-LPVVEANDYSFNDVTKTFEDAVSSPEEELSSEHETDCKPINFISIMIKQOKKHQIHK	761
D	b	2344	QLPRTSSPSTAS-----TKS-----SGSGKMSYS-----	2368
Q	y	762	VPTKQITASYOQYKNEQESRVTSDKVITPNAIOFKKFEVNVMSRRVYSPMDLNVSQF	821
D	b	2369	-PGRQL-----SQQNLKQGLGKSNMSSITPS-----	2394
Q	y	822	LPELSDSGFKDINFANYSNMNTNRRPSFTPL-STKNVLSINDNDPN-----VVEPP	871
D	b	2395	-----ESAKGLGNQNNNSNGSNMKVELSRMSTSTKSSQSESDRSPRALVNOGFIEKAP	2448
Q	y	872	EP--KSYAIIIRAARLANKAAPQAPPLPQROPQSSSTRGNSNKKRVSRFRVPTETIRRT	928
D	b	2449	SPTLRRRLLESASFESLS-----PSSRPDSTPTBSQAOTPV-----	2483
Q	y	929	SSALAPCDMYNDIFDDFGAGSKPTIKABGSKTLPSMDQDVKRLILNKKGVYQODEYINAK	968

Db	2484	---LSPS-----LPMOISLTHPSVQAGGRKLP-----PNLSPTLEYSDGRSRKSHDLAR	2530
QY	989	LYDQKPEKKNLSIVDPDEDRYEELQOTASINHNATIDSIYGRPDSITDMLPYLSDELKRRP	1048
Db	2531	SHSESPPSRL-----PVNRAGTWRKREHKSHTSSLPYSTRWRRRTGSSSIILASSE-----	2579
QY	1049	TALLSARLREDEQEVHRLBSNSVLYHPRGACATNSMLPPEPEELIN-----SPARNSNN	1104
Db	2580	-----SSERAKSEDEKH---VNSV---PGPRQMEKNQVPRKGTWRRIKESEISPTNTVSQT	2629
QY	1105	SDNVAISGNASTISFNOLDJNPFDDQATITOKIOEOPASKANTYVRGDDGLASAPERTP	1164
Db	2630	TSSAAGAGASKTLITYQMAAVRSYTEDVWRLEDCEPIN-----PRGSRG	2674
QY	1165	PTKR-----ESTISK---PAKUSSASPRKSPITIGSPVRVI---KNGSIAGIEPIPKA	1212
Db	2675	PTGNTPEVIDISIEKGNPISIKDSKDTQKQSVGSGSPVQYGLEMLRLNLSFIOVEAPBEDKG	2734
QY	1213	THHKRKKFCQGNELISNKKVDRDGG-----ISPSGSEHOOHNPMSWS--VPSQY----	1257
Db	2735	TE--TKAGQG---SPARVAVETGETCMAERTPSSSSSSKHSHTSSPCSTVAARVATPPNYNPSP	2789
QY	1258	----TDATSTVPDE	1267
Db	2790	RKSSADSTSARPSQ	2803

	RESULT 13
MOZ_HUMAN	ID
Q92794;	STANDARD:
15-JUL-1998 (Rel. 36, Created)	PRT; 2004 AA.
DT 15-JUL-1998 (Rel. 36, Last sequence update)	
DT 15-JUN-2002 (Rel. 41, Last annotation update)	
DE Monocytic leukemia zinc finger protein (zinc finger protein 220).	
GN ZNF220 OR MOZ.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX NCBI_TaxId=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=96376968; PubMed=8782817;	
RA Borrow J., Stanton V.P., Jr., Andresen J.M., Becher R., Behm F.G.,	
RA Chaganti R.S.K., Clavin C.I., Distèche C., Dube I., Frischauf A.M.,	
RA Horgan D., Mittleman F., Volinia S., Watson A.E., Hausman D.E.;	
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses	
RT RT a putative acetyltransferase to the CREB-binding protein.";	
RL Nat. Genet. 14:33-41(1996).	
CC -1- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.	
CC -1- SUBCELLULAR LOCATION: Nuclear.	
CC -1- DISEASE: PARTICIPATES IN A T(8;16)(P11:P13) CHROMOSOMAL	
TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMERA OBSERVED IN THE	
CC MM/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.	
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.	
CC -1- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC the European Bioinformatics Institute. There are no restrictions on	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
CC EMBL: U47742; AAC50662.1; -	
DR Genew; HONC:13013; ZNF220.	
DR MIM; 601408; -	
DR InterPro: IPRO01386; Histone_H1/H5.	
DR InterPro: IPRO02717; MOZ_SAS.	
DR InterPro: IPRO01965; ZnF_PHD.	
DR Pfam: PF00668; PHD; 2.	
DR Pfam: PF01853; MOZ_SAS; 1.	


```

QY 806 RRVSPMDLVLNVSQFLPELSEDSGFKDLNFANYSNNTN--RPRSTPLSTKNVLNIN 863
DB 1082 HALISPPSPSYNAE-----FEDC---DLNYSQVMSKGLFLESTSPNTS-----S 1124
QY 864 DPNVVEPEPPSYAEIRNARLSANKAPNOAPRLPEPOROSPTSNSSNKKVRSRFRVPTF 923
DB 1125 PPRCSPDPDR-----AEEIIMA--AAEKEAMLF---KGPVNYKKTVNSRIGKTSRARA 1172
QY 924 EIRRTSSALAPCDMYNDIFDDGAGSKPTI--KAGMKTLRPMDDVDVRIINAKGVNQ 981
DB 1173 QIKKSKAKLA-----NPSIVTKRKRRKQTKLVDDGKKKPRAKO-KTN 1215
QY 982 DEYIATKLVDOCKPRKNSTYTDPEDEYEELOQTASHINATIDS--IYGRPSDISTMLPY 1039
DB 1216 EKGTSRKHTTLTKDEKIKSGAGAEVFKLKNQVSEFASSGSQLFKKCD-----ML 1269
QY 1040 LSEDLKRP-----PTALLSADRL-----FME-----QEVHP-LRSNSVLV 1073
DB 1270 MGSADVHPLSASLPTGINKOQKLSGCFSSFLSKSKSVDLQTPSPSRHDLHPVAVCMSI-- 1327
QY 1074 HPGAGAA-----TNSMLPEPDELINSPARVNSNSDVA-----ISGNASTISFN 1120
DB 1328 --GPGVSKINORPHNOSAMFTLKESTLQKNIFDLSNHLISOVAAQNTQISSGSKIEDN 1385
QY 1121 QLDMPDQATIGOKIOEQPASKSANTVAGDDGLASABETPRTPTKESISSKRP----- 1175
DB 1386 ANNTIRNTLSIG-KLSEVRNS-----LESKIDQAVTPFLHCKXSOQOIVCI 1432
QY 1176 AKLSASPRKSPKIGSPVRIKKNGSIAGI-EP1PKATHKPKKSFQGNISNKKVRDGG 1234
DB 1433 AEOSSHSECSPGNTASEESQMPNCEVYLSRSPIKQIOMEOKRGFIDMSFKPE--R 1490
QY 1235 ISPSGSEHQHNPMS-----VSVPQOYDA-----TSTVP 1265
DB 1491 VAPRSLSEISOTKALSOCKKNRNPSTPSAFGQSGSLAVKELLOKROOKAONANTQDP 1550
QY 1266 DENKDVONKPREKOKKHNNHNNH-----HNHKOCTDIPGVVDELDIPVG 1310
DB 1551 LSNK---HQPKNKISGSLHNKANKRTRSVTSPRKRPRTRSKOKKIKRL--KYDSJN 1605
QY 1311 LOERGL-----FPRVLG1K-----NINLPDIT-----HKGR 1338
DB 1606 LONSSQLNVSVDSPRIFPSDGFESCYSLEDSLSPENHYNP-DINTIGTGFCSFYSGS 1664
QY 1339 FVLTDNGVHCVTPEYUNDDHNAIGKEFELTVADSLFEI-----LTKAS 1385
DB 1665 QVPAPADNL-----POKFLSD---AVODLFPQOALIEKNEFLSHDNOCDEKHHHTDSAS 1716
QY 1386 YEKPRGTL-VEYTEKKVYKS--RNRLSRLEFGSKDIITTKFVPLEVADTWANKFAPDGSFA 1443
DB 1717 WIR-SGTLSPLEIFEXSTIDSNENR-----HNQKKNSHPLTTKS 1755
QY 1444 KCVTD---LOQFEDOTGKASQFDLNCFNEMETMSNGNOPMKRGKPYKIAOLEVKMLYVP 1500
DB 1756 NSIMSFQVOQAD-----CLSEKSRLNRSYS----- 1783
QY 1501 RSDPRELLPTSRSAVESINELNNEQONFEGYLOHOGDCPIFKKRFKMGTSLLAS 1560
DB 1784 -----KEYFLSLPQPN--SDWIOGHTREKMGOSLDSANTSF---TALISSP 1825
QY 1561 EISHKTRAKINLSKYVDL-----IYVDKEN-----IDRSNHRNFSVDLLDHAERKIF 1608
DB 1826 D-----GELYDVACEDLEIVYSRNNMDLPTPDSRSTSSPSQSKNGSPTTPT 1874
QY 1609 ANGELIDFCAPKHKHKIWI--QNLOEITYYRNRFRROP 1644
DB 1875 ANILKPLMSPSRREIMATLHDLDSETIYOEPEFCSP 1912

```

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein HRX (All-1) (Fragment).
GN MLL OR HRX OR ALL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=C57BL/6J, and C57BL/6 X CBA; TISSUE=Spleen, and Lung;
RC MEDLINE=93317679; PubMed=8327517;
RA Ma O., Alder H., Nelson K.R., Chatterjee D., Gu Y., Nakamura T.,
RA Canani E., Croce C.M., Syracuse L.D., Buchberg A.M.,
RT "Analysis of the murine All-1 gene reveals conserved domains with
RT human All-1 and identifies a motif shared with DNA
RT methyltransferases."
RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
CC -! FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC MAY REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING
CC EMBRYONIC DEVELOPMENT.
CC -! SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -! SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -! SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -! SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -! SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC -! SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L17069; AA62593.1; -
DR MGD: MGI:96995; M11.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003889; FYrich_C.
DR InterPro: IPR003888; FYrich_N.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR001214; SET.
DR InterPro: IPR002857; ZnF_CXHC.
DR InterPro: IPR001965; ZnF_PHD.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR Pfam: PF02008; zf-CXHC; 1.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00542; FYRIC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50280; SET; 1.
DR PROSITE: PS01359; ZF_PHD_1; 3.
DR PROSITE: PS50016; ZF_PHD_2; 3.
KW DNA-binding; Nuclear protein; Zinc-finger; Metal-binding;
KW Transcription regulation; Alternating splicing; Polymorphism.
FT NON_TER 1
FT DNA_BIND 67
FT DNA_BIND 115
FT DNA_BIND 199
FT ZN_FING 1044
FT ZN_FING 1091
FT ZN_FING 1330
FT ZN_FING 1381
FT ZN_FING 1383
FT ZN_FING 1465
FT ZN_FING 1529
FT DOMAIN 1605
FT DOMAIN 1650
FT DOMAIN 3737
FT SET.

```

FT DOMAIN 35 41 POLY-GLY.
FT DOMAIN 459 469 PRO-RICH.
FT DOMAIN 1231 1238 POLY-PRO.
FT DOMAIN 3533 3536 POLY-GLU.
FT DOMAIN 3693 3697 POLY-GLU.
FT VARSPIC 1503 1505 MISSING (IN ISOFORM 2).
FT VARIANT 1497 1497 K -> T.
SQ SEQUENCE 3866 AA: 420976 MW: ADECS5E14E806FID CRC64;

Query Match 3.3% Score 285.5; DB: 1; Length 3866;
Best Local Similarity 19.6%; Pred. No. 0.00087;
Matches 384; Conservative 249; Mismatches 756; Indels 575; Gaps 94;

QY 13 KHSILOQPOSSASIFSPKPLNFPRTNPKSLDPNSSDYTSSEDOE-KGKEEKD 71
DB 2095 RHTSSLSPLRSKLRLM-SPV-----RTGSAYSRSSVSPSLCTATDPKASAKADRG 2147
72 TAFQTSFDRNFDLNSIDIOQTIOHQ-----QQPOQOOLSDT-----NNLDEFSEQ 121
DB 2148 GLSSSALHGHAPSSSSQRTVGSGSKTSHLDGSSPSSEVKCSALDLVPKSLVGEKNR 2207
QY 122 TPMTSTLDLTQKON---PTVDKVNENHAPTYINTSPNKSIMKATPKASPKKVAF---TV 174
DB 2208 TSSSKSTGDSASHATAYEIPRL---TPQVHNATPGELINIKISGFAPSPVSPSKDTV 2263
QY 175 TNPETHYPRNRVEEDOSQOKEDSVEP-----PLIOHQMKDPSQF 215
DB 2264 SYPLH--LRGQRSDRDHMDPSQSVKPSNDEGIKTLPGMGRPSILHEHIGSSSR 2321
QY 216 NYSDEDTNASTVPPTPLTKTPTFAQ--TLNKNNEVNSEPEALTLQMKLRNPSNLSDEK 274
DB 2332 DRQOKKSKSETCKEKKHSSSYLEPGQVTTGEGNLRP-----FADEV 2366
QY 275 VN---LYLSPNPNNSKNVSDMDSHLQNLQDASKNKTENIHLSFALKARKNDIENPLN 331
DB 2367 LTPGLGRPCNNVNSSEKIGD--KVLPLSGVPGQSTQ--VEGSKELQARKK----- 2415
QY 332 SLTNADIS-LSSGSSSLOSLRNDNVLESVQSPKKV-----NPG--- 373
DB 2416 -CSVKYVPLKMEGENSK-----NPKESGPGSPAHISVCPAEVVSASRSRPGAPG 2466
QY 374 -----LSLNDGKIGFSEVEVLSLPLDNRK----- 400
DB 2467 VQSPNNTLSODPOSNNQNLPRDORNLMTIDGPKPOEDGSFKRRYPRRSARAKNSMFG 2526
QY 401 -----LETTKEHAPREHNENFIIDAKSTNKGQLVSSDHLDSFDRS---YNTHQ 450
DB 2527 LTPLYGVSRYGEEDIPFYSNT---GKKRGRKSAAGQVDGADDLSTDEDDLYYNNFTR- 2582
QY 451 SILMLNLSASQSLANLEKOROTEOBOTOAPEEETSEFNKVKOEKPSMLEFVK 510
DB 2583 ---TVISSGGERLASHNL-----FREEQCDLPKISOLDGVDDI----- 2619
QY 511 VTIKEPVSAREIKA---PKR---EFSSRLIRIKNEDEIAPADI-----HPKKEAN 558
DB 2620 -TESTSTYATSRKSSQJPKRNGKENGTEMLKIDRPEDAGEKENVKSAVGHKNEPKLDN 2678
QY 559 SHVEDTALLKALNDESDTQNTSKMSIRFHIDSMKLEDS-----NDGREDND 611
DB 2679 CH---SVSRVKAQGGDSLEAQLSSLESSRVTSTPSPDKNLLDTYNACLLASDSDNNNSD 2735
QY 612 DISPREKDIANDV---SQTSDIIGDKGNSSSEITTKTLAPPRSDNNDKENSKLEPPA 668
DB 2736 DCGNLPDIDMDFLKNTPPSMQALGESPSSSELTTLGEGGLDLSNREK-----DIG 2788
QY 669 NNEISLOOLEVPHRKEDDSILANSNINAPREELTPVYEANDYSFND----- 716
DB 2789 LFEVPSQOL--PATEPVDSV--SSSISAEQFELPLLEPLDLSVLTTRSPVPSONSR 2844
QY 717 -----VTKTFDAYSFEES-----LSREH-----ETDSKP 741
DB 2845 LAVISDSGEKRVTTTEKSVASSEGDPLALLSGVDPAPRCHMTPDHFIGHMDADHISSPP 2904

QY 742 INFISIMWKOEKKKH-----QIHKVPRTKOILASVOQYANNEQESRYTSDKVKIPNAI- 793
DB 2905 CGSVBQHGNSODILTRNSGTGLOQVPSPTVYV-----QNGKYVPSSTDSPPSPQISNAV 2960
QY 794 -----QFKKEVNMVMSRRVSPMDLNVSOFLP-----LSEDSGRKDLNFA 837
DB 2961 QTTPHLKPATEKIL-----VNVNMOPRYLVQTLPLNGVTKOILTSPVSSPPSVMETNTS 3016
QY 838 -----NYSNNTNR--RSFTPLSTKNVLS-----NIDND 864
DB 3017 VLGRPGSSLLTTLGTLNPSLPSPSLFPAPSGLLSVPHNQHLSFPAAOSSFPFNISPP 3076
QY 865 PN-----VPEPERKSYAEIRNARR-LSANKAAPQAPRLPQROPSTRSNSKRVSRF 918
DB 3077 PSLGLIGVQPPDDPOLLSSEANQRTDLTTVATPSSG---LKKRIPSLHTRKKKLAPS 3133
QY 919 RVPTFEIRTSALAPCMYNDI-----PDDGAGAGKPIKKEGM-----KTLSSMKDD 968
DB 3134 SAP-----SNTAPSVNMTLLINFTPSOLSNHPSLLDLGSLNPSHHTVPI--- 3181
QY 969 VKRLNARKGYTODEYINAKLVQDKPKKNSIVTDPEDRYEELQOTASIHNTATIDSSYGR 1028
DB 3182 IKR--SKSGIMYEF-----QAP-----LLRP-----QSVGTATTAAGS----- 3213
QY 1029 PDSISTDMLPYLSDELKPPALLSADRLF---MEDVPHLRNSVYVHGAGATSS 1084
DB 3214 -STISOD---TSHLTSGPVALASGSSVVLNVWSQTTAAFTSTSY---PGHYTLANQR 3265
QY 1085 MLPEPDFE---LINSFARVNSNDNVVALISGNASTISFOLDNPFDDA---TIGOKI 1136
DB 3266 LIGTPDISISHLIKASHQSLGIDQDVALP--PSSGMFQOLGISQTPSAAMTAASI 3323
QY 1137 QEOPARSKANTVRGDDGLA-----SAP----- 1159
DB 3324 CVLPSSQTAAGMTAASPPEAEENHYKLGQNOQLAGKTGTLTSQRDRDPSAPQOPSIPT 3383
QY 1160 ETPRTPT-----KKESSSRPAKLSSASPPKSPKTISSPVRVLIKNSINGIERIPKAT 1213
DB 3384 QTAENPVGVSLEONKTLPSAKPA--SSASPGSSP-----SSGOQSGSSSVPGPT 3430
QY 1214 H-KPKKS-----FQNETSINKVRDGGISPSGSEHQOHNPSVSPQOYDATSTVDE 1267
DB 3431 KPKKAKRIQLRLDGKSVKKKIKVSHLRTSSSEAHNLRHRTDPA---PQSYTRPRRARE 3466
QY 1268 NKDYONKPREROKO-----KNNRNNNNHKKOTDIPGVVDEIDVGLQERK 1316
DB 3487 OODAGVQPSQKECGQAPAGVVALREVQATONPRANBOENAEPRKAMEBE--ESGFSSPLM 3544
QY 1317 LEFRVLGKINLNPDIINHKGR--FTLLIDNGVHCYTTPEYMMDHNAIGKFEFLTADS 1375
DB 3545 LMLQEOKRKESITERKPKKGLVPEISSDDG-----FOI-CAES 3582
QY 1376 LEFI---LTLKASYEKPRGTVEVEKRYVKSRN-----RLSRLFGSKDITTTK 1422
DB 3583 IEDAKKSLTDVQOERKSARLAKOLSFAGVNGLRMIGLIDHNAVLEIDBLAAKH-CRYK 3641
QY 1423 F---VPEVKDTWANKFAPDGSFARCYTIDQFEDQITGKASQEDLNCFNEMETWSNQ 1479
DB 3642 FRFHKPEEANRPLN---PHGS-ARAEVHLRK-----SAEDMFNF-----LASKHQ 3684
QY 1480 PMKRGKPKYIKOLEVKMLYVRSRDPRELLPTISAYSINELNNEQNNYFEGLLHDEGG 1539
DB 3685 PPEY--NPNDEEEEEVQLSARATSM-DLPMPR--FPHLKRTSKAEVGVRSPIHGG- 3739
QY 1540 DCPFFKRFKFLMGSLLAHS-----EISHTTRAKINLSKVVDLLYV---DKENIDRSNH 1591
DB 3740 ---LECKRNID-AGEMVLVEYGWNIIRSITQDKREKYYDSKIGCYMFRIDSEVVDATMH 3795
QY 1592 RNFSVLLLDHAFKIKFANGELIDFCAPNKHENKIMIONLOEII 1635
DB 3796 GNA-----RFINHS-----CEPNCTSRVITINDGQKHIV 3824

Mon Mar 17 12:50:24 2003

us-09-964-858-1.rsp

Page 21

Search completed: March 17, 2003, 12:24:53
Job time : 105.724 secs

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:21:44 ; Search time 25.9056 Seconds
(without alignments)
1889.932 Million cell updates/sec

Title: US-09-964-858-1

Perfect score: 8631
Sequence: 1 MNSYPSKLPIDKSHSLQLQ.....WNLMLQDQDQDQDQDSSSQ 1664

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8631	100.0	1664	US-09-599-652-2	Sequence 2, Appli
2	8631	100.0	1664	US-08-642-846-2	Sequence 2, Appli
3	8631	100.0	1664	US-09-264-604-2	Sequence 2, Appli
4	1211	14.0	236	US-09-599-652-3	Sequence 3, Appli
5	1211	14.0	236	US-08-642-846-3	Sequence 3, Appli
6	1211	14.0	236	US-09-264-604-3	Sequence 3, Appli
7	375	4.3	10182	US-09-134-001C-3159	Sequence 3159, Ap
8	348.5	4.0	3696	US-09-134-001C-5080	Sequence 5080, Ap
9	280	3.2	3788	US-09-336-447A-76	Sequence 76, Appl
10	274	3.2	1177	US-09-134-001C-5106	Sequence 5106, Ap
11	270.5	3.1	1863	US-09-099-753D-2	Sequence 2, Appli
12	270.5	3.1	1863	US-08-603-753D-2	Sequence 2, Appli
13	270.5	3.1	1863	US-08-986-106-2	Sequence 2, Appli
14	270.5	3.1	1863	US-07-007-678B-49	Sequence 49, Appl
15	270	3.1	2843	US-07-741-940-2	Sequence 2, Appli
16	270	3.1	2843	US-08-289-548A-2	Sequence 2, Appli
17	270	3.1	2843	US-08-452-654-2	Sequence 2, Appli
18	270	3.1	2843	US-08-370-235A-2	Sequence 2, Appli
19	270	3.1	2843	US-08-449-731-2	Sequence 2, Appli
20	269	3.1	2842	US-07-741-940-7	Sequence 7, Appli
21	269	3.1	2842	US-08-289-548A-7	Sequence 7, Appli
22	269	3.1	2842	US-08-452-654-7	Sequence 7, Appli
23	269	3.1	2842	US-08-449-731-7	Sequence 7, Appli
24	269	3.1	2843	US-08-452-655B-2	Sequence 2, Appli
25	269	3.1	2843	US-08-452-655B-7	Sequence 7, Appli
26	269	3.1	2843	US-08-450-582-2	Sequence 2, Appli
27	269	3.1	2843	US-08-450-582-7	Sequence 7, Appli

28	269	3.1	2973	US-08-821-355A-7	Sequence 7, Appli
29	269	3.1	2973	US-09-003-687A-7	Sequence 7, Appli
30	269	3.1	2973	US-09-136-605-7	Sequence 7, Appli
31	268.5	3.1	1535	US-08-755-587-185	Sequence 185, App
32	268.5	3.1	1852	US-08-425-061-24	Sequence 24, Appl
33	268.5	3.1	1852	US-08-825-886-24	Sequence 24, Appl
34	268.5	3.1	1863	US-08-425-061-16	Sequence 16, Appl
35	268.5	3.1	1863	US-08-480-784-2	Sequence 2, Appli
36	268.5	3.1	1863	US-08-483-553-2	Sequence 2, Appli
37	268.5	3.1	1863	US-08-487-002-2	Sequence 2, Appli
38	268.5	3.1	1863	US-08-483-554B-2	Sequence 2, Appli
39	268.5	3.1	1863	US-08-488-011B-2	Sequence 2, Appli
40	268.5	3.1	1863	US-08-825-886-16	Sequence 16, Appl
41	268.5	3.1	1863	US-08-850-727-2	Sequence 2, Appli
42	268.5	3.1	1863	PCT-US95-10202-2	Sequence 2, Appli
43	268.5	3.1	1863	PCT-US95-10203-2	Sequence 2, Appli
44	268.5	3.1	1863	PCT-US95-10220-2	Sequence 2, Appli
45	268.5	3.1	3969	US-08-061-376-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-599-652-2
; Sequence 2, Application US/09599652
; Patent No. RE37741
; GENERAL INFORMATION:
; APPLICANT: HOSTETTER, MARGARET K.
; APPLICANT: GALE, CHERYL A.
; APPLICANT: BENDEL, CATHERINE M.
; APPLICANT: TAO, NIAN-JUN
; APPLICANT: KENDRICK, KATHLEEN
; TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
; TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MUERING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
; STREET: 119 NORTH FOURTH STREET, SUITE 203
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/599,652
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/642,846
; FILING DATE: 03-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MUETING, ANN M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00280101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1664 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-599-652-2
; Query Match 100.0%; Score 8631; DB 1; Length 1664;
; Best Local Similarity 100.0%; Pred. No. 0;

REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1664 amino acids
type: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-846-2

Query Match 100.0%; Score 8631; DB 2; Length 1664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSTPEKLLPIDKSHLQLOPOSSSASIFNSPTKPLNFPRTNSKPSLDPNSSSDTYTSEQ 60
DB 1 MNSTPEKLLPIDKSHLQLOPOSSSASIFNSPTKPLNFPRTNSKPSLDPNSSSDTYTSEQ 60
QY 61 DEKGKEEKKDTAFQTSFDRNFDLNSIDIQFTIOHQOQOPOOQOOLSDTDNNLIDEFSF 120
DB 61 DEKGKEEKKDTAFQTSFDRNFDLNSIDIQFTIOHQOQOPOOQOOLSDTDNNLIDEFSF 120
QY 121 QTPMTSTLDLTQKNPTVDKYNENHAPTYINTSPNKSIMKKATPKASPKKVAFTVTNPEIH 180
DB 121 QTPMTSTLDLTQKNPTVDKYNENHAPTYINTSPNKSIMKKATPKASPKKVAFTVTNPEIH 180
QY 181 HYPDNFVEEDSOQKEDSEVERPLIOHOMKPSQFNYSDEDTNAVSPPTPLHTTKPTFA 240
DB 181 HYPDNFVEEDSOQKEDSEVERPLIOHOMKPSQFNYSDEDTNAVSPPTPLHTTKPTFA 240
QY 241 QLLNKNNEVSEPEALTDKKIKRENFSNLSLDKVLVLYSPNTNNNSKVNDSMDSHQNL 300
DB 241 QLLNKNNEVSEPEALTDKKIKRENFSNLSLDKVLVLYSPNTNNNSKVNDSMDSHQNL 300
QY 301 OPASKNKTENIHNLSFALKAPAKNDIENPLNSLTNADISLRSSGSSQSLQSLRPNRYL 360
DB 301 OPASKNKTENIHNLSFALKAPAKNDIENPLNSLTNADISLRSSGSSQSLQSLRPNRYL 360
QY 361 ESYVGSPPKVVNPGSLNDGIGKFSDEVESLLPRDISRKLETTKEDHAPENHNEFIDA 420
DB 361 ESYVGSPPKVVNPGSLNDGIGKFSDEVESLLPRDISRKLETTKEDHAPENHNEFIDA 420
QY 421 KSTNTKGGLLVSSDDHLSFDRSYNHTEOSILNLSASQOSISLNALEKQOTQOEQ 480
DB 421 KSTNTKGGLLVSSDDHLSFDRSYNHTEOSILNLSASQOSISLNALEKQOTQOEQ 480
QY 481 TQAAEEETSFSDNIKVKQEPKSNLEFYKVIKKEPVSAATEIKAPKREPSRILRIKNE 540
DB 481 TQAAEEETSFSDNIKVKQEPKSNLEFYKVIKKEPVSAATEIKAPKREPSRILRIKNE 540
QY 541 DELAEPADIHPPKENANSHVEDTDALLKALNDEESPTQNSTKMSIRFHIDSMKLE 600
DB 541 DELAEPADIHPPKENANSHVEDTDALLKALNDEESPTQNSTKMSIRFHIDSMKLE 600
QY 601 DSDNDGRENDDISREKSDIINDVSQTSIDIGDKYGNSSSETTTTTLAPRRDNDNDKEN 660
DB 601 DSDNDGRENDDISREKSDIINDVSQTSIDIGDKYGNSSSETTTTTLAPRRDNDNDKEN 660
QY 661 SKSLEDPANESLQOOLEVPHTKEDSDIANSNSINAPPELTLPVYEANDYSFENVYTK 720
DB 661 SKSLEDPANESLQOOLEVPHTKEDSDIANSNSINAPPELTLPVYEANDYSFENVYTK 720
QY 721 FDAYSSFEESLSREHETDSKPINFISIMWKQEKQKHQIHKVPTKQIILASYOQYKNEOS 780
DB 721 FDAYSSFEESLSREHETDSKPINFISIMWKQEKQKHQIHKVPTKQIILASYOQYKNEOS 780
QY 781 RYTSQVKIIPNAIQPKKKEVNVMSRRVYSPMDDLNVSOQLPELSEDSGFKLNPANYS 840
DB 781 RYTSQVKIIPNAIQPKKKEVNVMSRRVYSPMDDLNVSOQLPELSEDSGFKLNPANYS 840

QY 841 NNTNRPSTPLSTKNVLSINDPNVPEPEPKSYAEIRNARLSANKAAPNOAPPLP 900
DB 841 NNTNRPSTPLSTKNVLSINDPNVPEPEPKSYAEIRNARLSANKAAPNOAPPLP 900
QY 901 QROPSSSTRNSNKRVSFRVPTFEIRTSALAPCDMYNDIFDDPGAGSKPTTKAEQMT 960
DB 901 QROPSSSTRNSNKRVSFRVPTFEIRTSALAPCDMYNDIFDDPGAGSKPTTKAEQMT 960
QY 961 LPSMDKDDVKRIILNAKGVYQDEYINAKLVQDKPKKSIYTPDEDRKEELQOTASINAT 1020
DB 961 LPSMDKDDVKRIILNAKGVYQDEYINAKLVQDKPKKSIYTPDEDRKEELQOTASINAT 1020
QY 1021 IDSSIYGRPDSISTDMLPYLSDCLKPPTALLSADRLFMEOEVPYPLNSVLYHPGAA 1080
DB 1021 IDSSIYGRPDSISTDMLPYLSDCLKPPTALLSADRLFMEOEVPYPLNSVLYHPGAA 1080
QY 1081 TNSSMLEPDPFELINSPARNVSNNSDNVAISGNASTISFNQOLDMNFDDQATIGOKIOEP 1140
DB 1081 TNSSMLEPDPFELINSPARNVSNNSDNVAISGNASTISFNQOLDMNFDDQATIGOKIOEP 1140
QY 1141 ASKSANTVRGDDGLASAPETPTTKESISSKPAKLSASPRKSPITKGSFVRYLTKN 1200
DB 1141 ASKSANTVRGDDGLASAPETPTTKESISSKPAKLSASPRKSPITKGSFVRYLTKN 1200
QY 1201 GSIAGIEPIPKATNPKPKSFGQNEISNKKYRDGIGSPSSGSEHQOHPMSVSPQYTD 1260
DB 1201 GSIAGIEPIPKATNPKPKSFGQNEISNKKYRDGIGSPSSGSEHQOHPMSVSPQYTD 1260
QY 1261 TSTVPDENKDVQHKPREKQKHNHNNHNNHOKTDIPGVNDELPRDVLQERGLFFR 1320
DB 1261 TSTVPDENKDVQHKPREKQKHNHNNHNNHOKTDIPGVNDELPRDVLQERGLFFR 1320
QY 1321 VEGIKTNILPDTIHKGKFLTLIDNGYHCYTTREYIMDHNVAIGKEFELTVADSLFFIL 1380
DB 1321 VEGIKTNILPDTIHKGKFLTLIDNGYHCYTTREYIMDHNVAIGKEFELTVADSLFFIL 1380
QY 1381 TLKASYEKPRGLTVEYTERKVVKSRRNLSRLFGSKDITTTTKFVPVVKQTMANKRAPDG 1440
DB 1381 TLKASYEKPRGLTVEYTERKVVKSRRNLSRLFGSKDITTTTKFVPVVKQTMANKRAPDG 1440
QY 1441 SFARCYIIDLOQFEDQITGKASQFDLNCFNEMETMSNGQPKRGKPYKIAOLEVYKMLYVP 1500
DB 1441 SFARCYIIDLOQFEDQITGKASQFDLNCFNEMETMSNGQPKRGKPYKIAOLEVYKMLYVP 1500
QY 1501 RSDPREILPTISIRSAVESINELNNEQNNYREGYLHDEGGDCPIFKKRFELTMSLASH 1560
DB 1501 RSDPREILPTISIRSAVESINELNNEQNNYREGYLHDEGGDCPIFKKRFELTMSLASH 1560
QY 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNPSDVLLDHAFKIKFANGELIDPCAPN 1620
DB 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNPSDVLLDHAFKIKFANGELIDPCAPN 1620
QY 1621 KHEMKIWIQLOEIIYRNRRFPQWVLMLOQOQOQOQOQSSQO 1664
DB 1621 KHEMKIWIQLOEIIYRNRRFPQWVLMLOQOQOQOQOQSSQO 1664

RESULT 3
US-09-264-604-2
Sequence 2, Application US/09264604
Patent No. 6346411
GENERAL INFORMATION:
APPLICANT: HOSTETTER, MARGARET K.
APPLICANT: GALE, CHERYL A.
APPLICANT: BENDEL, CATHERINE M.
APPLICANT: TAO, NIAN-JUN
APPLICANT: KENDRICK, KATHLEEN
TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.

STREET: 119 NORTH FOURTH STREET, SUITE 203
 CITY: MINNEAPOLIS
 STATE: MINNESOTA
 COUNTRY: USA
 ZIP: 55401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/264, 604
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/642, 846
 FILING DATE: 03-MAY-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: MUETING, ANN M.
 REGISTRATION NUMBER: 33,977
 REFERENCE/DOCKET NUMBER: 110, 00280101
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-305-1217
 TELEFAX: 612-305-1228
 INFORMATION FOR SEQ. ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1664 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-264-604-2

Query Match 100.0%; Score 8631; DB 4; Length 1664;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSTPSKLLPIDKHSNQLQPOSSASIFNSPKPLNFPTNTSKPSLDPPSSSDTYTSEQ 60
 DB 1 MNSTPSKLLPIDKHSNQLQPOSSASIFNSPKPLNFPTNTSKPSLDPPSSSDTYTSEQ 60
 QY 61 DQKGEKKEDTAFQTSFDFNFDLNSIDIQITIQHQOQPOOQOOLSCOTDNLIDEFSE 120
 DB 61 DQKGEKKEDTAFQTSFDFNFDLNSIDIQITIQHQOQPOOQOOLSCOTDNLIDEFSE 120
 QY 121 QTPMTSTLDTKONPTVDKYNENHAPTYINTSPNKSIIMKATPKASPKKVAFTVTPEIH 180
 DB 121 QTPMTSTLDTKONPTVDKYNENHAPTYINTSPNKSIIMKATPKASPKKVAFTVTPEIH 180
 QY 181 HYDNNVEEEDSOQKEDSEVERPLIOHWKDPQOFNYSDEDTNAAVPPPTPLTTKPTFA 240
 DB 181 HYDNNVEEEDSOQKEDSEVERPLIOHWKDPQOFNYSDEDTNAAVPPPTPLTTKPTFA 240
 QY 241 QLLINKNNEVSEPEALTDMLKRENFSLSLDEKVNLYLSPPTNNNSKNVSDMSHLQNL 300
 DB 241 QLLINKNNEVSEPEALTDMLKRENFSLSLDEKVNLYLSPPTNNNSKNVSDMSHLQNL 300
 QY 301 QDASKNKTENIHNLSPALKAPKNDIENPLNSLTNADISLRSSGSSOSSQLSRNDNRVL 360
 DB 301 QDASKNKTENIHNLSPALKAPKNDIENPLNSLTNADISLRSSGSSOSSQLSRNDNRVL 360
 QY 361 ESVPSPKPVNPGSLNDGKGFSDDEVESLLPRDLSRDLKLETTKEHDAPEHNENFTDA 420
 DB 361 ESVPSPKPVNPGSLNDGKGFSDDEVESLLPRDLSRDLKLETTKEHDAPEHNENFTDA 420
 QY 421 KSTNTKKGQLVSSDDHLSFDRSYNHTEOSIINLNSASOSQISINALEKOKQOTQEOBQ 480
 DB 421 KSTNTKKGQLVSSDDHLSFDRSYNHTEOSIINLNSASOSQISINALEKOKQOTQEOBQ 480
 QY 481 TQAAPEEETSFSDNITKVQKPSNLEFVKVITIKKEPVSAATEIKAKREFSSRIILIKNE 540
 DB 481 TQAAPEEETSFSDNITKVQKPSNLEFVKVITIKKEPVSAATEIKAKREFSSRIILIKNE 540

QY 541 DEIAEPADIHPPKKNENANSHVEDTDALLKALNDEESDTQNTSKMSIRFHIDSDMKLE 600
 DB 541 DEIAEPADIHPPKKNENANSHVEDTDALLKALNDEESDTQNTSKMSIRFHIDSDMKLE 600
 QY 601 DSNDCGREDDNDSREKSDIILNDVQSDIITGDKGNSSSETTTTTLAPPSDNDKEN 660
 DB 601 DSNDCGREDDNDSREKSDIILNDVQSDIITGDKGNSSSETTTTTLAPPSDNDKEN 660
 QY 661 SKLEPPANNESIQOOLEVPHTEKSDSIILANSNINAPPELTLPVEANDYSEFNDVTKT 720
 DB 661 SKLEPPANNESIQOOLEVPHTEKSDSIILANSNINAPPELTLPVEANDYSEFNDVTKT 720
 QY 721 FDAYSFEESLSREHETDCKPINFISIMHKEKKHQHIVTKQIIASVOQYKNEQES 780
 DB 721 FDAYSFEESLSREHETDCKPINFISIMHKEKKHQHIVTKQIIASVOQYKNEQES 780
 QY 781 RVTSDDVKITPNAIQKFKFVNNVMSRRVVSPPDDDLNVSQFLPELSDESGFQDLNFAVS 840
 DB 781 RVTSDDVKITPNAIQKFKFVNNVMSRRVVSPPDDDLNVSQFLPELSDESGFQDLNFAVS 840
 QY 841 NNTNRPSFTPLSTKKNVLSNIDNDPVNVEPPEKSYAEIRNARLSANKAAPQAPPLPP 900
 DB 841 NNTNRPSFTPLSTKKNVLSNIDNDPVNVEPPEKSYAEIRNARLSANKAAPQAPPLPP 900
 QY 901 QROPSSSTRSNKRVSRFRVPTPEIRRTSALAPCDMYNDIEDPGAGAKPTTKAGMKT 960
 DB 901 QROPSSSTRSNKRVSRFRVPTPEIRRTSALAPCDMYNDIEDPGAGAKPTTKAGMKT 960
 QY 961 LPMDMDVYKRIINAKKGYTODEYINAKLVYDOKPKKNSIYTPDEDEYEELQOTASIHNT 1020
 DB 961 LPMDMDVYKRIINAKKGYTODEYINAKLVYDOKPKKNSIYTPDEDEYEELQOTASIHNT 1020
 QY 1021 IDSSIVGRPDSISTDMLPYLSDELKRPALLSADRLFMEQEVNPLRSNVLVHPGAGA 1080
 DB 1021 IDSSIVGRPDSISTDMLPYLSDELKRPALLSADRLFMEQEVNPLRSNVLVHPGAGA 1080
 QY 1081 TNSMMLPEPDFELINPARVSNNSDNVALISGAASITISFNQDMDNDADATIGOKIOEOP 1140
 DB 1081 TNSMMLPEPDFELINPARVSNNSDNVALISGAASITISFNQDMDNDADATIGOKIOEOP 1140
 QY 1141 ASKSANTVGRGDDGLASAPETPRTPKKESISSKPAKLSASAPRKSPKIGSPVRYIKKN 1200
 DB 1141 ASKSANTVGRGDDGLASAPETPRTPKKESISSKPAKLSASAPRKSPKIGSPVRYIKKN 1200
 QY 1201 GSTAGIEPIPKATHKPKKSFQGNELSNHNVYRDGSLSPSSGSEHQHNPMSVPSQYTDA 1260
 DB 1201 GSTAGIEPIPKATHKPKKSFQGNELSNHNVYRDGSLSPSSGSEHQHNPMSVPSQYTDA 1260
 QY 1261 TSTVPRDNKDVQKPREKQKNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNH 1320
 DB 1261 TSTVPRDNKDVQKPREKQKNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNH 1320
 QY 1321 VLGIRKININLPDINTHNGRTLTLDNGVHCVTTPREYMMDDHNAVIGKEFELTYAASLEFL 1380
 DB 1321 VLGIRKININLPDINTHNGRTLTLDNGVHCVTTPREYMMDDHNAVIGKEFELTYAASLEFL 1380
 QY 1381 TLKASYEKEPGLVEYTEKRVKSRNRLSRFGSKDIITTTKEVPTEVYKDTANKEFAPDG 1440
 DB 1381 TLKASYEKEPGLVEYTEKRVKSRNRLSRFGSKDIITTTKEVPTEVYKDTANKEFAPDG 1440
 QY 1441 SFARCTYIDQOPEDDQITGKASQFDLNCFMWETMSNGNPMKRGKPYKTAOLEVMKLYVP 1500
 DB 1441 SFARCTYIDQOPEDDQITGKASQFDLNCFMWETMSNGNPMKRGKPYKTAOLEVMKLYVP 1500
 QY 1501 RSDPREILPTIRSAYESINELNNEONNVEEGYLHDEGGDCYLFKKRPFKLMTSLIANS 1560
 DB 1501 RSDPREILPTIRSAYESINELNNEONNVEEGYLHDEGGDCYLFKKRPFKLMTSLIANS 1560
 QY 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVLLLDHAFKIKFANGELIDFCAPN 1620
 DB 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVLLLDHAFKIKFANGELIDFCAPN 1620
 QY 1621 KHEMKIWIQNLQETIYRNFRPQPVNMLLQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1664

Db 1621 KHEMKWIONLOEITVYRNFRROPVWNLMOOQOQOQOQOQSSQ 1664

RESULT 4
US-09-599-652-3

; Sequence 3, Application US/09599652
; Patent No. RE37741

GENERAL INFORMATION:

APPLICANT: HOSTETTER, MARGARET K.

APPLICANT: GALE, CHERYL A.

APPLICANT: BENDEL, CATHERINE M.

APPLICANT: TAO, NIAN-JUN

APPLICANT: KENDRICK, KATHLEEN

TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P. A.

STREET: 119 NORTH FOURTH STREET, SUITE 203

CITY: MINNEAPOLIS

STATE: MINNESOTA

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/599,652

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/642,846

FILING DATE: 03-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: MUEITING, ANN M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00280101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 236 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2

US-09-599-652-3

Query Match

Best Local Similarity 14.0%; Score 1211; DB 1; Length 236;
Pred. No. 5.5e-67;

Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SDEDINASVPPPTPLHTTKPTFAQLLNKKNVNSEPEALTDMLKRENFNSLSLDERKVL 277

DB 1 SDEDINASVPPPTPLHTTKPTFAQLLNKKNVNSEPEALTDMLKRENFNSLSLDERKVL 60

QY 278 YLSPTNNNNKSVSDMDSHLQNLQDASKKNTNENIHLNLSFALKAPKNDIENPLNSLTNAD 337

DB 61 YLSPTNNNNKSVSDMDSHLQNLQDASKKNTNENIHLNLSFALKAPKNDIENPLNSLTNAD 120

QY 338 ISLRSSGSSQSSLSQSLRNDNRVLESVPGSPKKVNPGLSLNDGJKGFSDEVESLPRDLS 397

DB 121 ISLRSSGSSQSSLSQSLRNDNRVLESVPGSPKKVNPGLSLNDGJKGFSDEVESLPRDLS 180

QY 398 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDDHLDSPDRSYNHTQSTIL 453

DB 181 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDDHLDSPDRSYNHTQSTIL 236

RESULT 5
US-08-642-846-3

; Sequence 3, Application US/08642846
; Patent No. 5886151

GENERAL INFORMATION:

APPLICANT: HOSTETTER, MARGARET K.

APPLICANT: GALE, CHERYL A.

APPLICANT: BENDEL, CATHERINE M.

APPLICANT: TAO, NIAN-JUN

APPLICANT: KENDRICK, KATHLEEN

TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P. A.

STREET: 119 NORTH FOURTH STREET, SUITE 203

CITY: MINNEAPOLIS

STATE: MINNESOTA

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/642,846

FILING DATE: 03-MAY-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: MUEITING, ANN M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00280101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 236 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2

US-08-642-846-3

Query Match

Best Local Similarity 14.0%; Score 1211; DB 2; Length 236;
Pred. No. 5.5e-67;

Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SDEDINASVPPPTPLHTTKPTFAQLLNKKNVNSEPEALTDMLKRENFNSLSLDERKVL 277

DB 1 SDEDINASVPPPTPLHTTKPTFAQLLNKKNVNSEPEALTDMLKRENFNSLSLDERKVL 60

QY 278 YLSPTNNNNKSVSDMDSHLQNLQDASKKNTNENIHLNLSFALKAPKNDIENPLNSLTNAD 337

DB 61 YLSPTNNNNKSVSDMDSHLQNLQDASKKNTNENIHLNLSFALKAPKNDIENPLNSLTNAD 120

QY 338 ISLRSSGSSQSSLSQSLRNDNRVLESVPGSPKKVNPGLSLNDGJKGFSDEVESLPRDLS 397

DB 121 ISLRSSGSSQSSLSQSLRNDNRVLESVPGSPKKVNPGLSLNDGJKGFSDEVESLPRDLS 180

QY 398 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDDHLDSPDRSYNHTQSTIL 453

DB 181 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDDHLDSPDRSYNHTQSTIL 236

RESULT 6

US-09-264-604-3

; Sequence 3, Application US/09264604

Patent No. 6346411
GENERAL INFORMATION:
APPLICANT: HOSTETTER, MARGARET K.
APPLICANT: GALE, CHERYL A.
APPLICANT: BENDEL, CATHERINE M.
APPLICANT: TAO, NIAN-JUN
APPLICANT: KENDRICK, KATHLEEN
TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETING, RAASCH, GERHARDT & SCHNAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/264,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 236 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
OTHER INFORMATION: amino acid positions 218-453 from SEQ ID NO:2
09-264-604-3
Query Match 14.0%; Score 1211; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 5,5e-67;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 SDEDTNNAVPPRPPLHTTKPTFAQLLNKKNNEVSEPEALTDMLKRENFSLDEKYNL 277
DB 1 SDEDTNNAVPPRPPLHTTKPTFAQLLNKKNNEVSEPEALTDMLKRENFSLDEKYNL 60
QY 278 YLSPNNNNNNKVVSDMDLQNLQDASKKNTNENTHNSLALKAPKNDIENPLNSLTNAD 337
DB 61 YLSPNNNNNNKVVSDMDLQNLQDASKKNTNENTHNSLALKAPKNDIENPLNSLTNAD 120
QY 338 ISLRSSGSSOSSLSQSLRNDNRVLESVPGSPKKVNPGLSLNDGKSGSDEVEVSESLPRDLS 397
DB 121 ISLRSSGSSOSSLSQSLRNDNRVLESVPGSPKKVNPGLSLNDGKSGSDEVEVSESLPRDLS 180
QY 388 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDDHDSFDRSYNHTEOSIL 453
DB 181 RDKLETTKEHDAPEHNNEFIDAKSTNTNKGQLLVSSDDHDSFDRSYNHTEOSIL 236

APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO: 3159
LENGTH: 10182
TYPE: PRN
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159
Query Match 4.3%; Score 375; DB 4; Length 10182;
Best Local Similarity 17.9%; Pred. No. 1.5e-13;
Matches 363; Conservative 326; Mismatches 688; Indels 650; Gaps 91;
QY 34 KPLNFPRTSKPSLDPNSSD--TYSSEODKKGKKEKKD-----TAFQTSFDRNFELD 85
DB 3424 EPLETAKNQLQANIDOKPSTDGTQOSVSYERKIQEAKDKINSINNVLANNPDVNAIRT 3483
QY 86 NSIDIQOTIQHQOQOPQ-----QQQLSQTDNNLIDFSFQTPMTSTDLTRQN----- 134
DB 3484 NKVEIEQ--INNELTQAKQGLTVDKQPLINAKTALQOSLDNOSTGTGMTATQVNAKRO 3542
QY 135 -----PTVDKYNENIAPRYINTSPRKSIIMKATKASKKKAFAFYTNTEIHNYDNRRE 188
DB 3543 KAEQYIQANAKKITEIENAPQSVQVSDSKVEQALSELNNAKSLARADQOELOAVNQLD 3602
QY 189 EEDSOQKEDSVPEPLIQHWKRPDSOFNSDEDTNNAVPPRPPLHTTKPTFAQLLNKNE 248
DB 3603 PTDLNKKKASTITAYNQRO-----QFSNELNSTKTITDITILEQNSVADVNNALNK 3655
QY 249 VNSEPEALTD---MKLRENFSLDEKYNLYLS--PTNNNSKNVSD----- 292
DB 3656 VREVOQKLNEARALLQNKEDNSALVRAKRLQOAVDVSTEGMTQOQFDVNSKQAAO 3715
QY 293 -----MDSHLQNLQDASKKNTN-----ENHNSLALKAPKNDIENPLNSLT--NAD 337
DB 3716 QETSKAQOVIDNGDATTTQOISNAKTNERALEALNNAKTGLRADKEEIQANVNLQNLND 3775
QY 338 -----ISLRSSGSSOSSLSQ-----LRNDNRVLESVPGSPKKVNPGLSLNDG 379
DB 3776 TSGKTPASIRKYNNAKSRIOQIDSAKKNANSILITNDNQSVQVTAALNKI 3826
QY 380 IGFSDVEVESLPRDLSRDKLETTKEHDAPEHNNEFIDAK----- 421
DB 3827 -----KAVQPELDRAIAMLK---NKE-----NNNALVQAKOQLOQVNEVDPTQGMTT 3871
QY 422 STNTNKGQLLVSSDDHDSFDRSYNHTEOSILNLSNASQOSLSINALEKOROTEOQ 478
DB 3872 DTANNYKSKKRAEDIEQAOQIINNCGATFEOQITNETNRVMO---ATNATKANKNDLRA 3928
QY 479 EOTOAEPPEETSFSQDNKIVKQEPKSNLEFVYVTKKEPVSAETIKAPRREKSSILRIK 538
DB 3929 DKSQI---ENAYNQLD-----NVD---TNGKKPASIQOYQAAQROALETOYNNAK 3972
QY 539 NE-----DELAEPAD-----IHPKKENANSHVEDTQALEKALND 574
DB 3973 SEAHQIENSNSVNEVQALQKVEAVOLKVNDAIHIIQNKENSALYATAKNOLOOSVND 4032
QY 575 D-----EESDTQNSNTKMSIRPHIDSKMLDSNODREDNDISFEKS--DIL 622
DB 4033 QPLTGMTOODSINNYEAKRNEAQSAIR---NNAEAVINNGDATAKOISDEKSVVEQALHL 4089
QY 623 NDVSO--TSDIIG-----DKYNSSEITTTKTLAPPSDNNKRENSKLE----- 665
DB 4090 NDAKOQITADTTELOTAVOQLNRGDTNNK-----KPRSTINAYNKAISQISLETOITSK 4142

666 DPAN-----NESLQO--OLEVPHTEKEDSILANSSNIAPEELTPVVEAND 710
1143 DANNAVIOKPIRTVOEYNNALQOVNOLQETAINOLOPLSNDA-----LKAARLNL 4196
711 YSSFDNVYTT-----FDAYSSFESLSREH-----ETD-----788
4197 ENKINTVOYTDGMOOSIEAYONAKRYAONESWTALALINNGDADEOQITETDRVNOQT 4256
739 -----SKPINFISIWKO--EKOK--KHQJHKVP-----TKOILASYOO--YKNE-----777
4257 TNLTOAINGLTY--NKEPFLAKTALONIDVOYSTGDMTOOSVANTIOQLAKNEINTI 4315
778 -----OESRVTSDKVKIPNAIQ-----FKKFK-----799
4316 NNVLANNPVNAIKTKAEERISNDLTQAKNNLOVDYTOPLKIKKQLODEIOTGTTDG 4375
800 -----EYVMSRRVSPDMDDL-----NVSOFLPEL-----825
4376 MPODSVDYNDLSAAIIEKGYNKLLKR--NPTVEQVESVANAQOVIDLOLONARTSLV 4433
826 -----SEDSGFKDLNFANYSNNTNPRSFPTPLSTKNVLSNIDNP 865
4434 PKOTLOEAKNNLENSINOOTDGTODSLNNYNKLAARQ-----NLEKISKVLGGOP 4489
Y 866 NYVEPEPEKSYAER--NARLSANKAAPNOA--PPLPPOQPSSTRSNSKRVSRFRPTP.923
4490 TV-----AEIRGNTDEANAHKQALDARSQTLNRPPIINHINNEHLNNAQKDNF 4540
924 EIRRTSSALAP-----CDMYNDLFDGAGSKPPIKAEKMTLPS--MDKDDVKR 971
4541 KAOVNS--APNHNTLETIKKAKADTLN--QSMTALSESJADYENOKOENYVLDASNNKR 4594
972 -----ILNKKKGV--TODEYINAKLVDOK--KKNSITDEPREDEELOQASTHNAT 1020
4595 QOYDANVNAKAGILNQTOSTPMASADYIDOKAEYVKRTKTALDGNORLEKAKQOALNHLNT 4654
1021 IDSSIIYGRDSDISTDMLPYLSDELKRPYALLSADRLFMEOEYVPLRSNSVLVHPGAA 1080
4655 LND-----LNDAROTLTDITN-----HSPRINSVNAKEKANT 4688
1081 TNSM-----LPEPEL-----INS--PARVNSNSDVAISGNASTISFQOLDMNF 1127
4689 VNTAMTOLKOTIANYDELHDNGYINADKDKDAYNNAVN-----NAKOL--INQOSANOA 4742
1128 --DOATIGOKIOEPASKANSYVGD-----DGLASAPETPRPTKESIS 1172
4743 QDDPRATINKYTORVNTTK--NDLNGMDKLALEAKRDANTTIDGLTYLNEAORAKA--KENVG 4799
1173 SKPAKLSSAPRKSPIKIGSPVYVLIKNGSIAGIEPTPKATHPKKSFQGNETISNKKVR 1232
4800 KASTKTNITSQLODYVQNLNIAQALRNS-----VNDVNNVKANS 4838
1233 GGSIPSSGE-----HOQHPSMV--SVPSQYTDATSTV-----1264
4839 NYINENGRKEAVYNOAVTHAQTLINAOSENPEMSRDVANOCTOAVNHAQNLHGOOKLEQA 4898
1265 --PDENKDOVOKHPRKOKOHNNHNNHNNHKKOKTDP-----GVVDEIDPDVG 1310
4899 QSSANTIEIGNLRLTYTOAKEKELYNSKOTREYVODLNOAKSLDSSMGTLSLVAKOP 4958
1311 LOERGKLFRRVLGIKNINLDPINTHKRGFTL-----TLDNGVHCYTPPE 1354
4959 TVOKTSVYL-----NEDQPEQASVNDSTJMGOTIINKTADPVLDTKYVONALISNTSTKE 5012
1355 YMMDDHNAVIGKEFELTVA--DSLEFLLIKASYEKPRGLIV-----TEKKVYKSRN 1406
5013 N-----ALHGEQKLLTAKTEAINALNTL--ADLNTPOKEAICTAINTAHTRTDVAEQS 5064
1407 RLSRLFGS-----KDIITTKFVPETEVKDWANFAPDGSFARCYID--LOQFEQOIJGKAS 1461
5065 KANOJISAMHTLRONISDNESVTNE-----SNYINAEPEKOHAFTELENNAKKATIVEDQA 5119
1462 QFDLNCFNE-----WETMSNGNOMPKRGRPYKIAOLEVKMLYVPRSPREILPTSIRS 1514

5120 TLDANSTINQAAQALITTKNMLDGEQJLRK-----ENAOEI-----5157
1515 AYESINELNNEONNVEGYLHQBEGDCPIFK-----KRFKMGTSILAHSEISHKTRAK 1569
5158 --MTLNQTLDAORSEKGLINSSQTRFEVASQLAKAKELNKVM-----EQLNHLING 5208
1570 ---INLSKYVDLTYVDKENIDRSNHRNPSVILLDDHFKFKFANGEL 1613
5209 NQMTNSSKFI-----NEDANOQOAVSNALASAEALKNSQNP 5247
RESULT 8
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080
Query Match 4.0%; Score 348.5; DB 4; Length 3696;
Best Local Similarity 18.2%; Pred. No. 1.6e-12;
Matches 363; Conservative 327; Mismatches 759; Indels 549; Gaps 90;
1 MNSPSPKLPIDKHSNQLQPOSSASIFNSPTKPLFPPTNSKPSLDPRSSSDTYTSEQ 60
1187 LNSTPHATQD--EKODALTRLQAKETAL-----NDINQAOITNQVDTALTSGIONTQ 1240
61 DOEKGEKKKDTAFOTSFDRNFIDLNSIDIQTOHQOQOPOOQOOLQOTDNILDEFSE 120
1241 VNNRKKQKQEAFTT-----NDI--VOQHQSQONNDATTEKEKYANNLVN--AS 1285
121 QTPWTSTLDTLKNQPNVYKVENHAPTYINTSPKSIKMKATP---KASPKVAVFTVTN 176
1286 QQWVYSKIDNATTNQIDGIVSDGROSINAITPDTSIKRAKKNDIDIKADKTI-----1339
177 PEIHHYDNNVEBEDSOQOKEDSVEPPLIOHOKKDPQOPVSDPDYAS-----VPPT 229
1340 -KIORINDATDEDEQENNR--IEEAKIEA--KDNIOERNSTRQVNEAKTNGINKIENI 1393
230 PRLHTTKPTPAQL--KNKN--VNSEPALTDK-----260
1394 TPATYVSEARQAVQKANKANOIHNQTPRATINEKEOALNRSASALARYOAOINAEHTT 1453
261 ---LKRENFSLNS-----LDEKVNLYLSPNTNNNSKNVSDMDSHL--QN 299
1454 QGVKTIKDAITSLSRINAQVVEKESARNAIEGKAQOQOFINND--NATDEKEVANN 1511
300 LODASKTKTENTHNSFALKAPKNDIENP-----LNSLTNADISLRSSGSSQSLQSLRN 355
1512 LVATKQKSLDNTINSLS-----SNNDVENAKVAGINELIAN-----1546
356 DNVLESVPSPKVPNGLSINDGIFGSEVAVESILPRDSRKLETTEKENDAP-----410
1547 -----VLPATVAKSKAKKDIOKLAQOI--NOIQTHQATTTTEKEKAIOLAN 1591
411 -----EHNNENFLDAKSTNTNKGOLLV---SSDHLDSFDRSYNHTEOSILN 454
1592 QKSNSEARTATONESHNSNGVQAQKSNGLHETELVMPDAHKKSDAKQSDINKYN--EGS--N 1647

QY 455 LLSNA-----SOSQISLNALE-----KORQTOEQOTQAEPEE-ETSESDNIKYQ 500
1648 TINTPDATDEEKQKALDLKIAKQAGYNKVDQAOQTNQOVSPAKTEALDITINIQANVAK 1707
QY 501 EPKSLLEF-VKVTIKKEVSAETETAPKREFSSRLIKNEDELEPADIHKKKEANS 559
Db 1708 KPSARVELDSKFEDKROKINIP-NATEEKODALORLNGKDEVKMLINDORNEVEQ 1766
QY 560 H-----VEDVDLALKKALNDESDPTQNSTKMSIRFHIDSDMLKDSNDGREDNDISR 615
1767 HKNIGLOLEFI---HAPTRKSDALO---ELQTFE-ISQFELINNNKDATHEEKDEAKR 1819
QY 616 ---FEKSDILNDV---SQTSDIIGDKYGNSSSEITF-----KTL 648
1820 LLEISKNTITINIQAOQTNQOVNADKNGMNEIATIPATITIKTDAKTAIDKKAQYTI 1879
QY 649 APPRSDNNDKENSKS-----LEDPANNEOLOQLEVPHTKEDSDILANSNAPPE 700
1880 INGNNDATDEEKAEARKLVEKAKITAKSITNSDTEREVNGAKTNG--LEKINNIOPTO 1937
QY 701 LTLPEVEANDYSFEND-----VTKTPEAYSSFE-----ESLSREHET 737
1938 ---TKTNNAQOETIDKQOEOLIQINNTPDATHEEKQOETNRVNNAGLAQALONINNAHST 1992
QY 738 D-----SKPINTSIWHKOEOK-----KHQIHVP 763
1993 QEVNESKTSNIATIKSVOPNVIKPTAISLQEVANNQKTLIGDNGNATDDEKEAKOLV 2052
QY 764 TKOITIASYQO-YKNQOESRVTSDK-----VKIPAIQFKKKEVMSRRVSPDMDL 816
2053 TOKLNEQOKIHESQDQOVNVAQOAITAIKLINAHKRODAINILF-NLASEKSDSI 2111
QY 817 NVSOFLPELSEDSGFKDLN-FANYSNNTNRPRESFTPLSTKNVLSINDPWNVEPPEK 874
2112 RANQOATTEKNTAIQSIDTDLAQRNNINGANT-NALVDENLEDQOKLOQIVLSTQK 2170
QY 875 SYAEIRNARKLSANKAANOAPPLPQOPQSPSTRSNSKRSRFRVPPFEIR-FRSSALA 933
2171 TOAKADIAQAIQOQSTIDQ-----NOVATTEKQOALERLQOETNGVNDRIQALAA 2222
QY 934 ---PCDMYNDIFDEFGAGSKPTI-----KAGSMKTLPSMDK--DVKRIINA 975
2223 NONVTEKNNILERT-RNVEPIVIVKPRANETIRKKAABOTLLINONODATILEEQIALG 2281
976 KKGVTODEYIN-----AKLVQDKPK-----KNSIVTDPEDRYEE 1009
2282 KLEEVKNALNQVSOAHSNNDVKIENNGIKAKISEVHPETIIRNAKOEIEQDAQSIDT 2341
QY 1010 L-----QOTASIHNAITIDSSITGRPDIS-----IDMLPYLDE 1043
2342 INANNKSTNEKSAIDVNNVAKIDAINNTNATTLQVLVNDAKNSGNTSISQILP-STV 2399
QY 1044 LKKPPTALLSARLEMEQEVHPLRSNVLVH-PCAGA-----ATN--SSMLPEPPELIN 1095
2400 VKTNALALASE-----AKKNALIIDOTPNATAEKEKANNAKVDRLQEADANILIK 2450
QY 1096 SPARNVSNNSUNVALISG-NASTISF-----NOLDNNFDOQATGOKIQOEPASKSA-- 1145
2451 AHTTDEVNINIKNOAVONINQAVEVIRKONVKNQLOCFIDQKKIEMTPDATLEEKAA 2510
QY 1146 ---NYVRGDDGLASAPETPRPTKESISSKPAKSSASAPKSPKIKSGPVAVIKN 1200
2511 NRLONVLTSTSDETLANVDH-----NEVDQALDKAP-----KIEIVAVQVSK 2555
QY 1201 GSIAGIEPIKATHPKKSFQGNESINHKVDRGDISPSSGSEHOQHNSVSPQYIDA 1260
2556 RDV--LNAIQAFNSQOIEQOENDEATNEKTEALN-----KINOLLNAKVNIDO----- 2604
QY 1261 TSTVPDENKDV-----QHKPREKOKHHHHHHHHHHHKKOKDI--PCGVNDE 1305
2605 ---AQSKVDVSAKTRSIQIDIEQOTPHROTKATGRHRLNEKANQOQSTIAHTHPSTIEE 2660

QY 1306 IPDVG--LOE-----RGK-----LFFRYLGIKNINLP-----DINTH 1335
Db 2661 ROEASAKLOEVLAKKAIAKIDKQOTNDVDYEKTVVNGIAETIENI-LPATVTKRKAADVNAE 2719
QY 1336 KGRFTLLDNGHCVCTTPPEYNNDDHNVAGKREFELTVADSLFELI--TLKASYEPKRGIL 1393
Db 2720 KEOKMLQINSDEAVT-----EKKLVASNLNHHVETINQOATEDAPDITNO 2764
QY 1394 VEYTEKK-----VKSRNRLSRFGS-----KDIITTFKVPTEVEDTANKFA 1437
Db 2765 VAVEKNKGIGTIRDIQPLVKKPFTAKSKIESAVEKKKTEINTQANTHDEVANE-GLNQLN 2823
QY 1438 PDGSFARCYIIDQOFEQDITGASQFDLNCFN-----EWETMSNGNOPMKRGKPYIAOLE 1493
Db 2824 QIHEKAKNDVNOSQTNQOVE-NAEONSIDQINNFRPDFSKRNAAVAELIYKAOQNKIDEIE 2882
QY 1494 VKM--LYVRSPRELPLPTSIRSAVESINELN--NEONNYFEGYLAHQEGDDPIKKRRF 1549
Db 2883 QEFSATQEEKNALQHLDBQVAYEITINSINQANTDNEVDNAKTSGLNNTETREPEYNNK-- 2940
QY 1550 KLMGTSLSLAHSEISHKTRAKINLSKVVDLIYDKENIDRSNHRNFSVDVLLDHAERKIFA 1609
Db 2941 ---KMAILKLYVSDTQEAIIIN-----GYPDATEDLEQANSKLNKILLKAKKQIGLA 2990
QY 1610 --NGELIDFCAPNKHMK 1625
Db 2991 HTNNEVDIYNEVSOKMK 3008
RESULT 9
US-09-336-447A-76
Sequence 76, Application US/09336447A
Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: ABEI, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACTYER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY:024
CURRENT APPLICATION NUMBER: US/09/336,447A
CURRENT FILING DATE: 1999-06-21
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 3788
TYPE: PRF
ORGANISM: Moraxella catarrhalis
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1036)..(3786)
OTHER INFORMATION: Xaa = any
US-09-336-447A-76
Query Match 3.2%; Score 280; DB 4; Length 3788;
Best Local Similarity 18.2%; Pred. No. 2.5e-08;
Matches 308; Conservative 285; Mismatches 700; Indels 396; Gaps 79;
QY 29 FNSPTKPLNFPRTNKKPSLDPRNSSSDYTSSEDDQKKEKKOTAFQSFDFNPLDLSI 88
Db 1132 YGSKSEQIDNPTEINLINLEARNAGKITKNAORNGSQIDNP--RTEINLINLEARNAV 1189
QY 89 --DIOOTIOHOQOOPOOQOOLSTODNNLIDFSFOTPMSTLDLKONP--TVDSKVN-- 142
Db 1190 IGDIGRKVSEQIDNPRTTEINLINLEARNMALEKNVEEGSEQID--NRTTEINLINLEARN 1245
QY 143 NHAPTYINTSPKSKIMKATPKASPKVAFTVTNDEIHHPDNRVEEEOQOKEDSVEP 202
Db 1246 NAXXANY-----ATPSTINALESNV-----EGGLXLSSEQIDNP 1279
QY 203 PLIOHOKDPSOFNYSDEDTNASVPTTPPLHTTKPTFAQLLNKNEVNSEPALITDMKLK 262


```
Db 1280 P-----RTEINLINEARNAKANYATPSIT-----INSALFENGSEQIDNPR-- 1321
Qy 263 RENFNSLIDEKYNLYLSPNNNSKNVSDMSHLONLODASKNKTNTENHNSLFALKAP 322
Db 1322 -----TEINLI-----NEARNASTIDLGKXVSEQIDNPREINLINEARNAKXAN 1365
Qy 323 KNDIENPLNSTIADISLRSSGSSQSLQSLRND-NRELVESGSPKPVPGSLNDGIR 381
Db 1366 YAPRSTINSITDLGTVDFGXXXXSEQIDNPREINLINEARNAKANYATPSITINSIT 1425
Qy 382 GRSDEVEVESLPRDLSRDKLETTKEHDAPEHNNEFDASTVTKNGQLLVSSDDLDSF 441
Db 1426 DLG-TIVDS---EQIDNPREINLINEA-----RNAVDALXTKYNALDXKVNSDXTSEQI 1476
Qy 442 DRSYNTTEOSIUNLNSASQS-----QISLALEKQRTQOEOQOAE-----PEERT 490
Db 1477 DNP-RTE---INLINEARNAKANYATPSITINSILAEOQLNGKTLTPVSEQIDNPRE- 1530
Qy 491 SFSDNIKVKOEKSNLEFVVTIKKEPVSAITEKAPKREPS-----SRIL 535
Db 1531 -----INLINEARNAKAKHDASTEKGMDSEQIDNPREINLINEARNAALLESVEGL 1585
Qy 536 RIKNEDEIAPADIHPKE-----NEANSVEDTDLAKKALNDESDTQNSTKM--SI 589
Db 1586 DLSGSEQID-----NPREINLINEARNAKNT--LIEKTANKSEQIDNPREINLINEA 1638
Qy 590 RPHIDSDMKLEDSNDQDREND---DISREKSDILNDVSOISDITGDGKNGSSSIT 645
Db 1639 RNAID--KNEYSIKSEQIDNPREINLINEARNASTIDLGTKSEQI---DNPREIN 1691
Qy 646 KTLAPPRSDNDKENSLEDP-----ANNESL-OOOLEY-----PHTKEDD 686
Db 1692 INARNAKNTLLEKSEQIDNPREINLINEARNAALHEQOLETLTSEQIDNPREIN 1751
Qy 687 SILANSSNAPPELTLPVYEANDYSSFNVDYTKFDAYSSFEESLSREHETS--KPI 744
Db 1752 INARNAKSSDSEQIDNPREINLINEARNAKNA--DADASF-ETLTSEQIDNPREIN 1809
Qy 745 ISIMHQ-----EKOKKHQIHKVPK-QILASQYKKNQOESR---VTSDKVIPIAIO 794
Db 1810 INARNAKATAIAKDSSEQIDNPREINLINEARNAKASENTQIAKSEQIDNPRE 1867
Qy 795 FKFEKVENYVSRVSHDMODLVNQSPLPELSE--DSGFDLNFPANYSNTNRPSTPL 852
Db 1868 -----EINL-----INARNAKLLDOKSEQIDNPREINLINEARNA--AATA 1910
Qy 853 STKANVLS-NIDNDPNVVEPEPEKSYAERNAKRLSANKAPN-QAPPLPPOPSSTRS 909
Db 1911 ITKNGKSEQIDN-----PREINLINEARNAKAKAXAANDRSQIDNPREINLIN 1961
Qy 910 NSKKRVSRFPVPTFEIRTSALAPCDMYDIDEDFGAGSKPTIKAGMKTLSMDKDV 969
Db 1962 EARNNAKANYATPSITINSNOADIA--ONQTDI-ODLAAYNE--LOSQIDN-PRTEIN 2015
Qy 970 KRLNKKGVTOEYIYAKVLDOCKPKKNSIVTPEDERYEELQOTASIHNAITDSSI 2029
Db 2016 NEARNAKQADIANINNTIYELAQOQDSQIDNPREINLINEARNAKANYATPSIT 2071
Qy 1030 DSISTDMLPYLDELKPPATALLSADRLFMEOEVHPLRSNVLVHPGAGATNSSM-- 1085
Db 2072 DALN-----SQIDNPREI--NLINAR-----NAIL-----GDTAIVSNGSQ 2111
Qy 1086 --LPEPDFELINSPARVNSNDVA-----ISGNASTISENOLDMMFDDO--ATIGOK 1135
Db 2112 IDNPREINLINE-ARNAKALLESVEGLDLDSGRSEQIDNPREINLINEARNAAL 2170
Qy 1136 IOBOPKSKANYRGDDGLASAPETPRPTKKESSISPAKLSASAPKSPIKIGSPVR 1195
Db 2171 VEEGLELSGRTI---DQSEQIDNPR-----EINLINEARNAKQAIANNIN 2216
Qy 1196 VIRKNGSIAGIEDLPKATHKPKSFGNGEISNKKVRDGGISPSGSGHQHNP----- 1250
```

```
Db 2217 XIYE-----LAQOQDOKS---EQIDNPREINLINEARNAKANYATPSITINQ 2262
Qy 1251 VSPVSOYTD-----ATSTVDENKDVQHKPREKOKHHNRH--HHHHHOKTIDPVD 1303
Db 2263 ADIAQNOTDIODLAAYNELQSEQID---NPREINLINEARNAKATHYNEKOTASQ 2319
Qy 1304 DEIPDVGLOERGLFEFVNLGKININLPDINTHKGRFTLTLDNGHCVTTPPEYNN--DDHN 1361
Db 2320 NPREIN-----INERNAKASENTQIAKSEQIDN-----PREINLINEARN 2365
Qy 1362 VAIGKEFELTVADSLFETLTK-----ASYEKPRGTVEYEKKVKSRRNLSRLGSKDI 1417
Db 2366 AMIGDPAI-VSNQDOKTKLKFYKSEQIDNPR-----TEINLI--NEARNAKGT 2415
Qy 1418 ITTKFVPT-----EVKPTWANKFAPDGSFACYIDLOQFEQITGKASQF 1463
Db 2416 PLDDXXPSSEQIDNPREINLINEARNAKANYATPS-----TTNSLHEQOLGKSEQI 2470
Qy 1464 D-----LNCFNEMETWMSNGQPMKRGKPYKIAOLEVMKLYVPRSDPREILPTSRSAVES 1518
Db 2471 DNPREINLINEARNAKANYATPSITINSILAEOQLNGKTLTPVSEQIDNPREIN 2521
Qy 1519 IN-----EINQONNYFEGYLHQEGDCPIFKKRFKLGTSLLAHSEISHKTRA 1568
Db 2522 RNAKANYATPSITINNNINNIYE-----LAQOQDQHSDDI 2556
Qy 1569 KINLSKYVD 1577
Db 2557 K-TLSEQID 2564
```

```
RESULT 10
US-09-134-001C-5106
; Sequence 5106, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5106
; LENGTH: 1177
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5106
```

Query Match 3.2%; Score 274; DB 4; Length 1177;
Best Local Similarity 19.6%; Pred. No. 1.2e-08;
Matches 218; Conservative 169; Mismatches 351; Indels 374; Gaps 54;

```
Qy 69 KKTAFQTSFDRNFDLNSIDIOQTIOHQOQDPQOQOOLSOTDNNLI-----DEFS----- 119
Db 3 KKGSVIMSWFDFKLGDDNGSN-DLLLRKNKRRSQSQSKONNODSLLPONNDIVSRPGRK 61
Qy 120 FOTPM-TSTIDLTQONTPVDCVVENHAPTYINTSPKNSIMKATPKKSPKVAATVTPNE 178
Db 62 FRPIQVSENYETOKN--ENYNEH-----NOBERTNDIMSYOH-----DNPE 102
Qy 179 I-----HHYDKNVEEDQS--QOKR-----DSVEPPLIOHMKDPQSFVNSDE 220
Db 103 FDSGSKHRRRRQAVSKHDQSKITQKQFADNNTNNNSV--FNQNDKSKKSSQGRKSIQ 159
Qy 221 DTNASVPTPLPLHTTKKPPFOALN--KNEVNSEPEALTDKLRNFNSNLSLDEKYNLY 278
Db 160 SENIKRANKMTKNTSTSEFTYLNHSFKSSEV--PSAIFGKKRRRPIENGVIAPPEHKL- 215
```

QY 279 LSPINNNSKNVSDMD--SHLONLQDASKNKTNENIHNLSFALKAPKNDIENPLNSLTNA 336
Db 216 -----NDKEIYQDEVSHSTKSIDASKVNSNDNNIE-----KNQ 251
QY 337 DISLRSSSSQSSLOSLNDNRVLESVPGSKVNPGLSLNGIKG----FSDEVESLL 392
Db 252 QKQOQTQOTESSSSNNHNN--VERSNVQTKRKTPNSKYDNTNINIENIVASQIVE--- 305
QY 393 PRDLSRDLKLETTKEHDAEHNENENFIDAKSTNTNGQLVSSDDHLD--SFDSY-----N 446
Db 306 -----EIRREKERVVLQKRRF--KKALQOKRQONQOSEDSDIQAIDEMAKQAOH 354
QY 447 HTEBSILNLNSASQSLSLALEKOROTOEOQTOAEPBEETSPSDNIKVKQ----EP 502
Db 355 YTGESSLDLENSNODS--SSNSLEKQ-----SNSNIDNKEKONNTP 395
QY 503 KSNLEFVVTIKKEPVSAETIKAPREFSSRLRIKNEDEIAPADIIHKKEENANSHE 562
Db 396 LFNTF-----ELDLDTTSDVYKVNNEETSK-- 421
QY 563 DTDALKKALNDEESDTTONSTKMSIRPHIDSDWKLIEDSN---DGDREDNDISREFK 618
Db 422 -----NDED-----LVSSNHYHSDDAEVEDAEYHELDNDRNQONSNS---Q 460
QY 619 SDILNDVQSDIIGDKYGNSSSETTTKTLPAPRSDNNDKENSKELEDPANNESLQOOLE 678
Db 461 DDILSSKSTSNM---YDMAIS-----ASVDNNTERAKSNED-----KNDTE 499
QY 679 VPH-----TKEDDSILANSSNIAPPELTLPVEANDVSSFNDDTKTFDAYSFEESSLS 732
Db 500 ITHLDGTTSAKVSDKISNTNHNLEFQDKNVKLNKVNLSKSNSDT-----GQTR 549
QY 733 REHETDSKPIFISIMHKOQKQKHQHK-----VPTKQIIASVQOYKNEQESRVTSD 785
Db 550 KQREGSGSRPFN-VLMTPSDKKRMQDNHKKVSVPPELKEPKQANMHRKQSEKNS---- 603
QY 786 KVKIPNAQOFKKFKRVNVMSRRVSPDM-----DLVNS-----QFLPEL 825
Db 604 -----EEFKQIN--NRETDSNYSIESNGIEHDMSSSDEHAYETPSKQOODOIOL 652
QY 826 SEDSGFKDLNFANYSNNTNRPSEFTPLSTKNVLSNIDNPNVPEPPEKSYAEIRNARL 885
Db 653 ODDFHEFENAHAKI--NNSN-----ETGNOINISHSKR-- 683
QY 886 SANKAAPQAPPLPQROPSSSTRSNSKRVSR---FRVPTFEIRRTSSALAPCDMYNDIF 942
Db 684 --SOYSTNESKNIDQTSNNSSTSNQNFQIRKGPNIKLPYOL--LEAPEPHERQDQWI 738
QY 943 DD-----FGAGSKPTIKAGCKMTLPSPMDK-----DDVKRTL 973
Db 739 DNKKOELMDALYFNVPAEAKNVTEG---PSVTRFELSVKGVVSRITLALODDIKMAL 794
QY 974 NAK-----KGYTODEYINAKLVQDKPKR--NSIYTPEDRYEELQOTAS--IHNATI 1021
Db 795 AAKDIRIAPIGTSLVGIEVNPQNPVKVNLISIESPKFKNTESKLTYAMQYRINNE-- 852
QY 1022 DSSIYGRPDISTDMLPYLSDELKAPPALLS 1053
Db 853 -----PLIMD-IAKTPHALIA 867

RESULT 11
US-08-603-753D-2
Sequence 2, Application US/08603753D
Patent No. 5891857

GENERAL INFORMATION:
APPLICANT: HOLT, JEFFREY T.
APPLICANT: JENSEN, ROY A.
APPLICANT: PAGE, DAVID L.
APPLICANT: KING, MARY-CLAIRE
APPLICANT: SZABO, CSILLA I.
APPLICANT: JETTON, THOMAS L.
APPLICANT: ROBINSON-BENION, CHERYL L.

APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESS: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: ductal carcinoma in situ, invasive
CELL LINE: breast cancer and normal breast tissue
CELL LINE: not derived from a cell line
ORGANELLE: no
IMMEDIATE SOURCE:
LIBRARY: cDNA library derived from human
CLONE: obtained using published sequence
POSITION IN GENOME:
CHROMOSOME/SEGMENT: unknown
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: BRCA1 protein
LOCATION: 1 to 1863
IDENTIFICATION METHOD: observation of mRNA and
IDENTIFICATION METHOD: antisense inhibition of BRCA1 gene
OTHER INFORMATION: BRCA1 protein has a negative
OTHER INFORMATION: regulatory effect on growth of human mammary cells.
PUBLICATION INFORMATION:
AUTHORS: MIKI, Y., et. al.
TITLE: A strong candidate gene for the breast and
TITLE: ovarian cancer susceptibility gene BRCA1.
JOURNAL: Science
VOLUME: 266
PAGES: 66-71
DATE: 1994

RELEVANT RESIDUES IN SEQ ID NO: 2: grainin box
 RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 1214-1223
 us-08-603-753D-2

Query Match 3.1%; Score 270.5; DB 2; Length 1863;
 Best Local Similarity 19.1%; Pred. No. 3.6e-08;
 Matches 339; Conservative 239; Mismatches 610; Indels 591; Gaps 83;

18 QLOPSSASIFNSPTKPLNPRRTSKPLSDNSSD-----TYSSEOQEKKEKKPT 72
 Db 170 RIQPKTSYI-----ELGSDSEDIYKNAKYGSGVQELLQITPQGT 212
 Qy 73 AFQTFDR-----NEDLNSIDIOOTIQH-----OOQFOOOLSOQTN 112
 Db 213 REIILDSAKKACER-----SEIDVINTENHOPSNNDLNTEKRAARHREKQG--SSVN 268
 Qy 113 NLIDFESFOTPM-----STLDLTKONPTVDK--NENHAPTYINTSPNKSIMKAT- 162
 Db 269 LHEVECGTTHASLQHESSILLTFKDRMNVKEAEFCNKSQGLARSOHNRWAGSKETC 328
 Qy 163 -PKASPKKAFVTVNPEIHHPDNRYEEDQSOQ-----EDSVETPLQ----- 206
 Db 329 NDRRTPS-----TEKKVDLADPLCEEREMNKQKLPCEBNPDTEDVETLNLNSIQV 382
 Y 207 HOMKDPSONYSD-----DINASVPPTPLHTTKPTFAQLLNKNVNSEPE 254
 Db 383 NEM-----FSRDELLGSDSHDESESNKV-----ADVLDVLEW----- 419
 Qy 225 ALTKMLKRENFSLDERKVLVLSPTNNNSKNVSDMSHLQNLQDASKRNTENIHN 314
 Db 420 -----DEYSGSS--EKIDLAS-----DPH-----EALICK-SDRVHS 449
 Qy 315 LSFALKAPKNDIENPLNSLTNADILSRSSGSSQSLQSLRNDKRVLESVPGRKKVNPGL 374
 Db 450 -----KSVESDIEDKI-----FKITYRKASLSPNLSHVENI----- 482
 Qy 375 SLNDGKGFSE-----VVESILPRDLSRDLKLETKEDHAPENENITDAKSTNKGOLL 431
 Db 483 -----IGAVSEPOITIOERPLTNKLRKRPTSGILHP-----EDT-----KKAOLA 524
 Qy 432 VSSDHLDFDRSYNTEOS--ILNLINSASQSLMALEKQ-----OTQEOEOT--Q 482
 Db 525 VQKTEPM--INQTNQTEONGQVMTNNGHENKTKGDSIQNEKNPNPLESLEKESAFKT 582
 Qy 483 AAE-----EETSFSQNIKKQEPKSN-----LEPVKTIKPEPSATE 522
 Db 583 KAEPISSISINLELIMHNSKA--PKNRLRKSSTRHIALELV-VSRNLSPPNCTE 638
 Qy 523 IKAPKEFSRILRIKNE-----EIAEPADIHPK--KENEANSHEVDIAL 568
 Db 639 LQIDSCSSSEELIKKKYKNQPMVHNSRNQLQMEKKEPATGAKKNKNEOTSKRHSDTTP 698
 Qy 569 KALANDESDPTONSTKNSIR-----HISDMKLED--SNDQREDNDI-----SR 615
 Db 699 ELKLTNAPSFTKCSNTSELKEFVNPSLPREEKEKLETVKYSNNA--EDPKLIMLSGER 756
 Qy 616 FEKSDILNVSQTDILGKYG--NSSSETTKTLAPPSGDND-----KENSKL-- 664
 Db 757 VLQTEKSVSSSISLVPGLDYGIOESISLELVSTLQAKTEPKVCYSQACAFENPGLIH 816
 Qy 665 ---EDPAN--ESLQOL--EVPHTKEDDSLANSNNIAPPELTLPLVEANDYSFN-- 715
 Db 817 GSKDRNRDETEGFKYPLGHEVNHNSRFTSIEMSESLDAQVLQNTFVYSKQSGAPSPNG 876
 Qy 716 ---DVTKTDAVSSFEESISREHETSKPLNFIISMHKQKKHQ--HKVPTKQIIAS 770
 Db 877 NAEBCATPSAHSG--SLKKQ--SPKVT--ECEQKEENQCKNSINIKPVQTVNITAG 928
 Qy 771 YQGYKNEOSRYTSDKVKIIPNALQFKKFEVNNMSRVVSPDDDLNVSGF--LPELSEDS 829
 Db 929 F--PVVGQKRPVDNAKCSIKGSRFCLSSQFNGNETGLTIPNKHGLQNLQRYRILPPIK 987
 Qy 830 GF-----KDL--NFANYS-----NNTNRPRESFTPLSTKVLNIDNDPVWVEPPEP 873

Db 988 SEVTKCKKNLLEENFEHSHMSPEREMNGENIPSTVSTISRNNIRBNVKE----- 1038
 Qy 874 KSYAELRNARRLSANKAARQAPLPQROPSSSTRNSKNKRVSRFVPEIRTRTSALA 933
 Db 1039 -----ASSNINEVGS-----STNEVGSSINEG 1062
 Qy 934 PCDMYDIFDDFGAGSKPTIKAGMKTLPSMDKDVRIILMAKGVTOEYINAKLVQK 993
 Db 1063 SSD--ENIQAEIQRNNGPKLNA-----MLRLGLQPEVYKQSLPGSN 1102
 Qy 994 PKKNSIVTPDEKRYEELQOTASIHNATIDSYGRPDSITDMLPYL--SDELKKPPTALL 1052
 Db 1103 CKRPEI--KQOEYEVQGT-----VNTDPSPLYISNLQEP----- 1136
 Qy 1053 SADRLMEQDEYHPLRSNVLVHNGAQAATNSSMLPREDELINSPPARVNSNDVAISG 1112
 Db 1137 -----MGSSHAGVCSFPPDIL-----DDEIKR 1161
 Qy 1113 NASTISFNOLDMNFDDQATIGOKIOEQPASKS-----ANTVRGDDGLA 1156
 Db 1162 DT---SFAENDIK--ESSAVFSKVQKGLSRSPFTHTHLAGYRGAKKLESSEBLS 1217
 Qy 1157 SAEPTPTPTKKEKISSKRAKLSSASPRKSPKIKGSPVRVYKKNKSGIAGIEPIPKATHKP 1216
 Db 1218 S--EDELPCQFHLLFGKVNNIPSQTRHSTV-----ATECLSKNTEEN 1259
 Qy 1217 KKSFGQ--NEISNAKYRQDGISPSGSEHOQHNPMSVNP--SQYTDATSTVPENKDVCH 1273
 Db 1260 LSLKNSLNDSCSNQVILAKASQEHNLSEETKCSASFSSQCSLEDELDTANTNTQDPFLIG 1319
 Qy 1274 KPEKQKQKNNHNNHNNHNNQ--KTDIPGVYDELIPDVGLQERKGLFFVLGIKNINLP 1330
 Db 1320 SSKQMR-----HOSQGVGLSKDELVPDDEBGRGTELENNQ----- 1356
 Qy 1331 DINTHKGRTLLDNGHCVTPEYNNMDH--NVAIGKEPELTVADSLFILTAKASYEK 1388
 Db 1357 -----EESQMSNLGEAASGCESETSVSD----- 1381
 Qy 1389 PGCTLEVEYTKKVVKSRNRLSRFGSKDITTTKRVTEVYKQDPMANKFADGSPARCYID 1448
 Db 1382 -----CSGLSOSDILT-----TOORDMQUHL--4-----IK 1406
 Qy 1449 LOOFEOITGKASQFDLNCENEMETMNGNQPM-----KRGRPYKIAOLEVMILYVPS 1502
 Db 1407 LQDEMALEAVLQHGSPNSSTPILSDSSALEDLRNPQSTSEKYLQTSQSSSEFPIS 1466
 Qy 1503 -DPREILPTIRSAYESINELNNEONNYFEGYLHQEGDCPIFKKRFKIMGTSLAHSE 1561
 Db 1467 QNEGXSADKFEVSADSTSKKE-----PGEVRSFSPKCPSLDDRMYHMSGGSLQNRN 1521
 Qy 1562 ISHKTAKTILSKVVDLIYVDKENIDRNSNHRNSDVL 1600
 Db 1522 YPQOE--ELIKYVD--VEEQULESGPHDLTETSYL 1553

RESULT 12
 us-09-099-753-2
 ; Sequence 2, Application US/09099753
 ; Patent No. 6149903
 ; GENERAL INFORMATION:
 ; APPLICANT: HOLT, JEFFREY T.
 ; APPLICANT: JENSEN, ROY A.
 ; APPLICANT: PAGE, DAVID L.
 ; APPLICANT: KING, MARY-CLAIRE
 ; APPLICANT: SZABO, STILLA I.
 ; APPLICANT: JETTON, THOMAS L.
 ; APPLICANT: ROBINSON-BENION, CHERYL L.
 ; APPLICANT: THOMPSON, MARILYN E.
 ; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
 ; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
 ; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
 ; NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch, 800 KB storage
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: ductal carcinoma in situ, invasive
CELL LINE: not derived from a cell line
ORGANELLE: no
IMMEDIATE SOURCE:
LIBRARY: cDNA library derived from human
CLONE: obtained using published sequence
POSITION IN GENOME:
CHROMOSOME/SEGMENT: unknown
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: BRCA1 protein
LOCATION: 1 to 1863
IDENTIFICATION METHOD: observation of mRNA and
OTHER INFORMATION: BRCA1 protein has a negative
PUBLICATION INFORMATION: regulatory effect on growth of human mammary cells.
AUTHORS: Miki, Y., et. al.
TITLE: A strong candidate gene for the breast and
TITLE: ovarian cancer susceptibility gene BRCA1.
JOURNAL: Science
VOLUME: 266
PAGES: 66-71
DATE: 1994
RELEVANT RESIDUES IN SEQ ID NO: 2: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 1214-1223
US-09-099-753-2

Query Match 3.1%, Score 270.5; DB 4; Length 1863;
Best Local Similarity 19.1%; Pred. No. 3,6e-08;
Matches 339; Conservative 239; Mismatches 610; Indels 591; Gaps 83;
18 OLPOSSASIFNSPTKPLNPPRINSKPSLDNSSD-----TYSSEDOEKKEKKOT 72
170 RIQPKTSYVI-----ELGSDSSSEDTVNKATYCSVDOELQITTPGT 212
73 AFOTSFDR-----NFDLNSIDIOQTIOH-----QOPOOQOQOISQDN 112
213 RDEISLDAKKACEF-----SETDVTNTEHQPSNDLITTEKRAERPEKYQG--SSVSN 268
113 NLIDEFSFQTEPT-----STLDLTKQNPYDKV---NENHAPYINTSPKSIKKAT- 162
269 LHVEPCGTNTHASSLOHENSILLTKDHMNEKAEFCNKSQPGIARSGHNRWAGSKETC 328
163 -PKASPKVAFVTPNPELHHYDNRVEEDOSQK-----EDVEPPLIQ----- 206
329 NDRRTPS-----TEKKVDLADPLCEKKEWKKOKLPKSENPREDVEDVPWITLNSIQV 382
207 HQMKDPSQFNYSDE-----DTNASVPPPTPLHTTKPTPAQLLNKNEVNSPE 254
383 NEM-----FSRDELLGSDSDHDESESNKY-----ADVLDVLENY----- 419
255 ALTDKLRKRENFSLDEKYNLYLPTNNNSKNVSDMSHLQMLDASKNKTNENIHN 314
420 -----DEYSGSS--EKIDLLAS-----DPH-----BALICK--SDRVHS 449
315 LSFALKAKKNDIENPLNLTNADISLRSSGSSQSLQSLRNDNRVLESVPGPKVNGCL 374
450 -----KSVESDIEDKI-----FGKYRKKAISPLNLSHVENLI----- 482
375 SLNDGKGFSPD---VVESELLPRDLSRDKLETTKEHDAPEHNNENFIDAKSTNTKQGL 431
483 -----IGAIVSPQIIOGRPLNKLKRRRRTSGLHP-----EDFI-----KKAOLA 524
432 VSSDHLDSFDRSYNHTEOS--ILNLNSASOSQISLNALEKQ-----QTOEOEQT--Q 482
525 VQKTPEM--INGTQOTEDONGVNMNITNSGHNKTKGDSIONEKPNPIESLEKESAKRT 582
483 AAE-----EETSFSDNIRVKQEPKSN-----LPEVYVTKKEVSATE 522
583 KAEPILSSISNLELNLIMNSKA---PKNRLRRSSRTHIALLY--VSRLLSPNCTE 638
533 IKAPKREFSSRLIRIKNED-----EIAEPADIHPK--KENANSIHVEDTDL 568
639 LQIDCSSESEELIKKKYQNMVPRHSRNQLMGKREPARGAKSKNKPNOTSKRHSDFRP 698
569 KALNDEBDSDTTONSTKMSIRF-----HIDSMKLED--SNDGDRDNDI-----SR 615
699 ELKLTNAGSGFTKCSNTSELKEFVNPSLPREKEKEKLETVKVSNNNA--EDPKDLMSGER 756
616 FEKSDIINDVSGTSDIIDKYG--NSSEIITTKTLAPRSDND-----KENSLSL-- 664
757 VLQTERSVESISSILVPTGDTGTQESISLLEVSTLGAKKTEPKCVSCAFAENKGLIH 816
665 ---EDPANV-ESLQOOL---EVPHTKEDSILANSSNIAAPPELTLPVVEANDYSSEN--- 715
817 GCSKRNNDTEGFKYPLGHEVNHRSRETISEMESLDAQYLQNTTKVSKROSFAFSMPG 876
716 ---DVTKTFDAYSSFEESI-SREHETDSKPIINFISIMHOKQOKHO--IKHYPTKOIIAS 770
877 NAEECATFSAHSG---SLKQO---SPKVPF--ECEQKEENGKNESNIKPVQTVNITAG 928
771 YQKYNEQESKRYTSKVKVIPAIALQFKKREVVWMSRRVYSPMDLNLNSQF--LPLSEDS 829
929 F--PVYGQKDPVADNAKCSIKGSRFLCSSQFERGNETGLTPKHGLLQNLNPKRIPLPLPK 987
830 GF-----KDL---NFANVS-----NNTNRPSTPLSTKNVNLSDNDPNVVEPEP 873
988 SFVTKCKKNLLEENFEHSHSPEREKNGENIPSTVSTISRNKINENYFKE----- 1038


```

Db 483 -----IGAFVSEPOIIQERPLTKLKRKRPTGLHP-----EDFL-----KRAOLA 524
Qy 432 VSSDDHLDSFDSRYNHTEOS--ILNLNSASQOISLNALEKOR-----OTOEOEOT--Q 482
Db 525 VOKTPREM--INQOTNOEQGYMNITNSGHEKTKGDSIQNKPNPIESLEKESAFKT 582
Qy 483 AAEF-----EETFSFSDNIKYOEPKSN-----LEPVYTIKPEVSATE 522
Db 583 KAEPISSISNELLEIMHNSKA---PKNRLRKRKSTRHIALELY--VSRNLSPPNCTE 638
Qy 523 IKAPKREPSRILRIKNEO-----ELAEPADIHPK--KENEANSHEEDDALL 568
Db 639 LQIDSCSSSEELIKKKYNOHPVHRSNRLQMEKSEEPATGAKKSNKPREQSKRHDSDTFP 698
Qy 569 KRALNDEESDTQNSTKMSIRF-----HIDSWKLED--SNDGDREDNDI-----SR 615
Db 699 ELKLTNAPSFTGCSNTSELKEFVNPSPREKEKELEYKVSNN--EDPKMLMSGER 726
Qy 616 FEKSDILNDVSDTIDGKYG--NSSSETTKTLPAPRSDND-----KENSLSL-- 664
Db 757 VQTERSVSSSSISLVPGDYGTQESISLLEVSTLGAKTEPKVCVSCAFAENPKGLIH 816
Qy 665 ---EDBANN--ESLQOOL--EVRPTKEDDSTLANSNIAPEEELTLVPEANDVSSFN--- 715
Db 817 GSKORNDTEGKYPYLGHEVNHRSRTSLEMESELDQOYLQTEKYSKQSPAPSPNG 876
Qy 716 ---DYTKTFDVASSFESISREHETDSKPIFISIMHKOEQKHO--IHKVPYTKOIIAS 770
Db 877 NAEBCATFSAHSG---SLKQK---SPKYTF--ECQEKENQCKNSNIKPVQTVITAG 928
Qy 771 YQOYKKEQESRYTSQVKYLPNAIQFKKFEVNVMSRRVYSPMDLNVSOE--LPEISEDS 829
Db 929 F--PVYQKQKPVNMACSIKSGSRFCLSSQFRGNETGLITPNKHGILLQNYRILPFPPIK 987
Qy 830 GF-----FDL---NFANFS-----NMTNRPSTPLSTKVLNSINDPVPVPPER 873
Db 988 SFPKTKCKNLEENFEESMSPREKGNENIPSTVSTISRNIRENVKE----- 1038
Qy 874 KSYAETIRNARLSANKAANQAPRPPROPSPSTRSNKRVSRFVPTFEIRRTSSALA 933
Db 1039 -----ASSNINEVGS---STNEVGSISNEIG 1062
Qy 934 PCDMYNDIDDDGAGSKPTIKAGMTTLPSMDKDYKRIINAKKYQODEYINAKLYDOK 993
Db 1063 SSD--ENIQAEIGRNGPLNA-----MLRGLVLOPEVYKQSLPQSN 1102
Qy 994 PKNSIVTDPEDRYEELQOTASTIHNTIDSSIYGRPOSISTOMLPLYL--SDELKKPPTALL 1052
Db 1103 CKHPEL---KKOYELEVOT-----VNTDSPYLIIDNLEOP----- 1136
Qy 1053 SADRLEMEQEVHPLRSNSVLYHPGAGATNSSMLPEPDELINSAPARVSNNSDNYAISG 1112
Db 1137 -----MGSSHASQVCSFETPDLL-----DDEEIKNE 1161
Qy 1113 NASTISFNOLNMFDOATIGQIQOQRPASKS-----ANTVAGDDGGLA 1156
Db 1162 DT---SFAENDIK--ESSAIVESKSVOKGELSRSPPTHTHLAGOYRRGAKKLESSEENIS 1217
Qy 1157 SAPETPRPTKRESISSKPAKLSSAPRKSPIKIGSPAVYIKKNGSIAGIEPIPKATHPK 1216
Db 1218 S--EDELPCFQHLRGKANNIPSOSTRHSTV-----ATECLSKNTEEN 1259
Qy 1217 KRSFQO--NEISNHVARDGISPSSGSEHQOHNPMVSVY--SOYTDATSTVPDENNDVOH 1273
Db 1260 LSLKNSLSDSCNOVILAKASQDEHLSSEFKCASJLSSQCSSELELTLANTNQDFFLIG 1319
Qy 1274 KREKQKQKHHHHHHHHHKK--KTIDIGVVDDELIDYVGLQERGLKFRVGLGIKINILP 1330
Db 1320 SSKOMR-----HQSQSGVGLSDKELVSDDEERGGLLEENNO----- 1356
Qy 1331 DINTHKGRTLTLDNGVHCVTTPEVNMDDH--NVALGKFEFLTVASLEFILTLKASYEK 1388

```

```

Db 1357 -----EEOQMSDNLGEAASCESETSVSD----- 1381
Qy 1389 PRGLIVEETKVKYKSRNRLSRFGSKDITTTKFPVTEKQDWMANKFAPDGSFARCYID 1448
Db 1382 -----CSGLSSOSDILT-----TOORDMOINL-----IK 1406
Qy 1449 LQOFEQITGKASQFDLNCFNENETWSNGQPY-----KRGKPYIAOLEVLMYVPRS 1502
Db 1407 LQOEMAELEAVLQHOCSQSPNSYSPIISDSSALEDLRNPQOSTSEKYLQTSQKSEYPI 1466
Qy 1503 -DREILPISIRAYESINLNEQNNYFEGYLOHQBQCPPIKRRFKMLGTSLLAHSE 1561
Db 1467 QNPEGSADKFEVSAUSSTSKNE-----PGVERSSPKCPSLDDRWYMHSGSLQNN 1521
Qy 1562 ISHKTAKINLSKYVDLIYDKENIDRSNHRNPSVILL 1600
Db 1522 YPQOE---ELIKYVD---VEBOQLBESGHDLTETSYL 1553

RESULT 14
US-09-007-678B-49
; Sequence 49, Application US/09007678B
; Patent No. 6342483
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: OBERMILLER, PATRICE S.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: METHOD FOR DETECTION AND TREATMENT OF BREAST CANCER
; FILE REFERENCE: Attorney Docket No. 6342483 1442-1-2-2
; CURRENT APPLICATION NUMBER: US/09/007,678B
; PRIOR FILING DATE: 1998-01-15
; PRIOR APPLICATION NUMBER: 08/373,799
; PRIOR FILING DATE: 1995-01-17
; PRIOR APPLICATION NUMBER: 08/182,961
; PRIOR FILING DATE: 1994-01-14
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Microsoft Wordpad
; SEQ ID NO 49
; LENGTH: 1863
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1472)
; OTHER INFORMATION: Xaa-any amino acid
US-09-007-678B-49

Query Match 3.1%; Score 270.5; DB 4; Length 1863;
Best Local Similarity 19.1%; Pred. No. 3,6e-08;
Matches 339; Conservative 239; Mismatches 610; Indels 591; Gaps 83;

Qy 18 QLOPOSSASIFNSPTKPLNFPTNSKPSLDPNSSD-----TYSQEOQEKKEKKT 72
Db 170 RIQPOKTSYVI-----ELGSDSSDPTVKNKATYCSVGQDELQITPOCT 212
Qy 73 APQTSFDR-----NFDLNSIDIQOTIOH-----QOQOQOQOQOTON 112
Db 213 RDEISLDSAKKACER-----SETDYVTEHHQPSNNDLNTTEKRAAERHPEKTOG--SSVSN 268
Qy 113 NLIDFSPQTPMT-----STDLTKQNPVTYKV--NENHAPTYINTSPKNSIMKAT-- 162
Db 269 LHYEPCGTMTNTHASSILOHENSLLTTRDMNVKEAKFCNKSQDGLARSGHNMWAGSKETC 328
Qy 163 -PRASPKKAAFTVYTNPEIHHYDPNRYEEDQSQK-----EDSVYEPPLIO----- 206
Db 329 NDRRPPS-----TEKKVLDLMDPCLCEKKEWVKQKLPCCSENPRDTEDPVWITLNSIQKY 382
Qy 207 HOKKQPSQFNYSDE-----DTNAYVPPRPPLHTTTPTFQOLLKNKNVNSSE 254
Db 383 NEW-----FSRDELLGSDSDSHGSESNAKY-----ADVLIDVLENY----- 419

```

QY	255	ALUDMKAKREBENSLUDEKVENLYLSTPTNNNNKSNVSDMDSHLOQLDQASKNKTENTENHN	314
Db	420	-----DEYGGSS--EKIDILLAS-----DPH-----BALICK--SDRVHS	449
QY	315	LSFALPKAKNDINPLNLSLTLNADISLRRSGSSOSSLOSRLNRNRYLESVPGSPKPVNFG	374
Db	450	-----KSVESIEDIKI-----FKYIRKKASLPLNSHTYENLI-----	482
QY	375	SLANDGKIGFSDE--VVESLLPRDLSROKLETTKEHADAEHNNENFIDAKSTNTKCOLL	431
Db	483	-----IGAFVSEPOIIIOERPLTKLKRKRRTPSGLHP-----EDFI-----KKADLA	524
QY	432	VSSDHLHDSDRBYNTBOS--ILNLNLSASOSQISLNALEKOR-----QYOEBOQT--Q	482
Db	525	VOKTPRE--INQSTOTEOGOVYMTNTNGHENKTKDSTIQNEKPNPFIESTLEKSAAKT	582
QY	483	AAEP-----EETSFDNIKVKOEPPSN-----LBEVYTIKKREVSATE	522
Db	583	KAEPISSISINLELNMHNSKA--PKNRLRKSSSTHIALELY--VSRRLSPNCTE	638
QY	523	IKAKREFFSRILIRKNE-----ETAEPADIRPK--KENANSHVEDTDALL	568
Db	639	LQIDSCSSSEIKKKKYNQMPVHSHNLQIMGKEPATGAKKSNKPNQTSKRHSDOFF	698
QY	569	KKALNDEESDTONSTKMSIR-----HDSMDKLED--SNDGREDDNDT---SR	615
Db	699	ELKLTJNAPGFTKCSNTSELKEFVNPSLPREKEEKLFFVKYSNNA--EDPRDMLSGER	756
QY	616	FEKSDILNDVQSODSIIDGKYG--NSSSEITTKTLAPRSDND-----KENSXL--	664
Db	757	VLOTREKVESSSISLVPGDYGTQESISLLENTSLTAKATEBNKCVSQAFAENKGLIH	816
QY	665	---EDPAN--ESLQOOL--EVPHTKEDDSILANSSNIAPEBELTLPVYEANDYSSFN--	715
Db	817	GCSKDNNDTEGFKPYPLGHEVHNSRBTSTJEMESELDADYLQNTFKYRSQSFAPFSMPG	876
QY	716	---DYTTPLAYSSFEELSREHETOSKPINPLSTIHKOEOKKXHO--IHKVPTKQIITAS	770
Db	877	NAEECATFSAHSG--SLKKO--SPVYTF--ECQKEENOGKNESNKPQYVNTIATG	928
QY	771	YQCKNKOESRYVSDVKYIKNALIOFKFKFEVNVMSRRVYSPMDLDVNSQF--LPELSDS	829
Db	929	F-PVYGOKKQPVDMAKCSIKGSRFELSQFRENETGLTTPKHHGLONPYRIPLFPFK	987
QY	830	GF-----KDL--NFANYS-----NNTNRPRSFPTLSTKNVLSINDNDPNVVEPER	873
Db	988	SFVKTKCKKMLBENFEHSMSPEREMGNBNIPSTVTSIRNNIRENVKE-----	1038
QY	874	KSYAETINARRLSANKAAPQAPPLPPQOPSPSTNSNSKRYSRRRVFPFETIRTSALA	933
Db	1039	-----ASSSININEVS-----STNEVGSSINEIG	1062
QY	934	PCDMYNDIPDDPGAGSKPTIKAGMKTLPSMOKDVKRLNKKKGVTODEYINAKLVQK	993
Db	1063	SSD--ENIQIELGRNRCPKINA-----MLULVQLEPYVKOSLPGSN	1102
QY	994	PKKNSITDDEDRYEELQOATSIHNTATIDSIYGRPDSISTMLPYL--SDELCKPPATLL	1052
Db	1103	CKHPEI---KKQYEVEVQI-----VNTDFSYLLSDNIEQI-----	1136
QY	1053	SADRLMEQEVHPLRSNSVLYHFGAGATNNSMLPEPDFELINSBARNVSNNSDNVAISG	1112
Db	1137	-----MSSHAGQCSBTPDDL-----DDGEIKE	1161
QY	1113	NASTISFNQDAMFDDOATIGQIQOQPAKS-----ANTYKRGDDGLA	1156
Db	1162	DT---SFAENDIK--ESSAYFSKVSQKGELSRSPSPETHHLAQYRGARKKLESSEEMLS	1217
QY	1157	SAPETPTPKKKSISSKPAKLSSAPRKSPIKIGSPVAVIKKNSINGIETPIKATHPK	1216
Db	1218	S--EDELPCFOHLLGCKVANNITPSOSTRSHV-----ATBELSKNTEEN	1259

```

QY 1217 KKSPOG--NFISSNHKVRADGIPSSSCSEHOJNNHVSVP--SOYUATSTVDEKDKQJH 1273
Db 1260 LLSLAKNSLNCSSQVILLAKASOPHNLSSETKCSASLSSQCELEDLTANNTQDFRLG 1319
QY 1274 KPEKOKKHHNNHHNNHHNNKQ--KTJDIPGVYDDELIDVGLQERKLEFFRVLGIKINLP 1330
Db 1320 SS:KQMR-----HQESQGVGLSDKELYSDERGTEGLENNO----- 1356
QY 1331 DINTHKRGFTLLDNGVHCYTPPEYMDQH--NVAIGKFEFTLVADSLFILLTKASYEK 1388
Db 1357 -----EEOQMDSNLGEAAGSCSESTVSIED----- 1381
QY 1389 PGCTIVEVEKTKVYKSNRLSRLFGSKDILTTTKFVPEVQCTYANKFARDGSEFARCYID 1448
Db 1382 -----CSGISSSODILT-----TOORITMOHNL-----TK 1406
QY 1449 LQGFEDDITGKASOFDLCNEMETMSNNOQM-----KRKRPKIAQLQLEKMLYUPRS 1502
Db 1407 LQGFMALEAVLIDHQHSPNSYPSRILSISSALEDLRNPQESTSEKVLQTSQKSEYIIS 1466
QY 1503 -DPEILLPTIRSAYESINELNNEQNNNYEGYIHDGDCPLFFKKRFRKLMGTSLAHSE 1561
Db 1467 QNPEGASADKEFVESADSSYTSKNE-----PQVERSSPSCPSLIDRWYMHSCSGSLQMRN 1521
QY 1562 ISHKTRAKINTLSKVYDLYDKNDIDNRSHNRNSDULL 1600
Db 1522 YPPOE--ELIKYVD--VEQDLEESGPHDITTEYSTYL 1553

```

RESULT 15
 US-07-741-940-2
 Sequence 2, Application US/07741940
 Patent No. 5352775
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOF.
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch, Mckie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/741,940
 FILING DATE: 19920109
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.035574
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO. 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids

TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-741-940-2

Query Match 3.1%; Score 270; DB 1; Length 2843;

Best Local Similarity 19.1%; Pred. No. 7e-08; Mismatches 287; Conservative 216; Indels 428; Gaps 64;

QY 22 QSSASISFNSP---TKPLAFPRNTSKP-----SLDPSSSDYTSRODEKKG 66
Db 924 RSSAHTHTNTYNTFKSENSNTCSMPYAKLEYKRSSNDLSVSSNDGY-GKRGAKMS 982
QY 67 EEKKDTAFQTSF-----DRNFDDNSIDIOQTIOHQO---QO 100
983 ISYSDEDSKFCSYGQYPADLAHKIHSANHMDNDGELDTPIYSLKYSDEQLNGROS 1042
QY 101 PQOOOOLSTQDNNLIDFESFQTPMTSTLDLTQONPTVDKVNENHA-PTYINTSPKSIK 159
1043 PSQNERWAPKHIIEDEI-----KSEQROSRNOSTYTPVTESTDDKHL-- 1087
QY 160 KATPKASPKVAFTYVNPFIHHYPDNRV-----EEEDSOOKEDSVE---PPLIQHOM 209
Db 1088 KQPHFGQOECSPTIRSKANGSETNRVGSNHGINOVQSILQEDDYEDDKPTNYSERY 1147
QY 210 KDPQFNVSDEDTNASVPTPLHTTKPTFAQLLNKNNEVSEPEALTDMKLKRENFSL 269
Db 1148 SEEDHEEERPTNYSI-----KYNEEKRHVDQPIDYSLKYA--TDI 1187
QY 270 SLDEKYNLTLSPNNNSKNVSDMDSHLONLQDASKNTNEN-IHLSALKA--PKNDI 326
Db 1188 PSSQKOSFSSKSSGQSKTEHMSSSSENTSTPSSNAKROHLPSSAQSNGQOKAA 1247
QY 327 ENPLNSLTNADISLR-----SSGSSQSSLOSIRND--NRVLESYPGSPKKVNPGL 374
Db 1248 TCKVSSINQETIQTCVEDTPICFSRCSLSSLSAEDIEGNOQTQEAANDNTLIQAI 1307
QY 375 SLNDGKGFSDVEY---SLPRDISRDKLETKKHAPEHNNENF-IDAKSTNTNGQO 429
Db 1308 KGRIGTSAEDPVSEVPASQHPRTKSSRLQSSLSSESARHKAVFEPGAKSPSKSGAO 1367
QY 430 LTVSSDDH-----LDSFRSRYNHTQESILNLNLNASSQO-----ISLN 467
1368 TRKSPREHVVQETPLMFESRCTSVSSLDSE-----SRISASSVOSEPCSGWGISPS 1421
Db 468 AL-EKQQRQTOEOQOQAEEETSEFSDNIKYQEPKSMLEFYKVIKKEPVSAEIKAP 526
1422 DLPDSCGQTPMPSRSKTPPEPQTA---QTKREVPKN-----KAPTAEKRESGP 1467
QY 527 KR-EFSSRLRLK-----NEDETAEPADIIHKKE 554
Db 1468 KOAAVNAAVQRYVLPDADTLHFATESTPDGFSQSSSLALSLEDPFOKDELRIIMP 1527
QY 555 NEANSHVEDTDLKKALNDEESDTTONSTKMSTRFHIDSDMKLEDSMDGREDNDIS 614
Db 1528 VOENDNGNETSEQKESNEQEKAEKT-----IDSEKDLDDS---DDDDIE 1573
QY 615 REEKSDILNDVSQSDIIGDKYGNSSSEITTKTLAPPRSDNNDKENSKELEDPPANNESIQ 674
Db 1574 ILEEC-IISAMPTKSRGKKKPAQTASKL-----PPVAKRPSQLPYVKLLPSQNR-LQ 1625
QY 675 QCLEVPHRTKEDP-----SILANSSNIAPPEEL----- 701
Db 1626 POKHVSFTPGDDMPRYVCVEGTPINFTATSLSDLTIESPPNELAEGVGAGGAGSEFE 1685
QY 702 ---TLPVYANAYSSFNQVTKTEDAASSFEESLSREHETDSKPINFISIMHKOEKOKKHQ 758
Db 1686 KRDITPTBERSTDEAGGKTSV-TIPELDDNKAEEGDTLAECLN-----SAMPKGSNR 1739
QY 759 IHKVPTKQIIASVQO-----YKNE---QESRVTSDEKVKIPNAIQKFKFEVNVMSR- 806
Db 1740 PPRV--KRTMDVOQASASSASPNNKQDLGKKKKPTSPVKPIQNTETRYTRVKKNADSKN 1797

QY 807 ----RVVSPMDLNLVSOFLPELSEDSGFKDLNFANYSNNTNR-----PRSFTEL 852
Db 1798 NLNAERVFSDNKD--SKQNLKNNSKD--FND---KLPPNEDRYGSAFSDPHHYPTI 1849
QY 853 -----STKNVLSINDNPNVVEPEPEPKSYAEIRAKRLSANKAAPNPAPLPQORQS 905
Db 1850 EGTPEYCFSSKNDSLSLDFDDVDLSREK--AEIRAKENKSEAKVYTSHTL-----T 1901
QY 906 STRSNSNKRVRFRVPTPEIRRTSALAPCDMYNDIFDFGAGSPPTIKAEQMKTLPSMD 965
Db 1902 SNOQSANKTQALAKQ---INR-----GQPKPILQO--STFPOSS 1937
QY 966 KQDVKRIILNAKKGVYODEYINAKLVQDKP-----KKNSTV--TD 1002
Db 1938 KDIPDR-----GAATDEKIQNFATIENTPVCFSSHNSLSLSDIDQENNNKENEPKETE 1991
QY 1003 PEDRYE-----LQQTASHNNTIFSSITGRDSTIDMLPL 1040
Db 1992 PDSQGEPSKPOASGYAPKSFVHEDTPVCFSSRNSLSLSDISE---DDLQECISSA 2046
QY 1041 SDELKPPPTALLSADRLFMEOEVHPLRSNSVLVHPGAGAAATNSSMLPEPDELINSPARN 1100
Db 2047 MPKKKKP-----SKLKQDNKHSRNNKGIL--GEDTLTLQKIQRDSEHGLSPD-- 2095
QY 1101 VSNNSDNVAISGNASTISFNQLDMNFDAQATIGQKIQEQPASKSANTYVRGDDGLASAP 1160
Db 2096 -SENFDMKAIQEGANSI-----VSSLHQAAAACLSRQ--ASSDSDSILSKSGISLQSP 2147
QY 1161 TPTPTKRESISSKPAKJASSAPRSPKPIKIGSPVRYIKKNGSIAGIEPIPKATHPKKSF 1220
Db 2148 FILTPDOEE---KPE-TSNKGR-----ILKPEKSTLE--TKIESESKTI 2188
QY 1221 QGNEISNHRVYRDGIGSPSSGSEHQHNP-----SNVSPSOYTDATSTVPDEN 1268
Db 2189 KQAKKYKSLTIGKVSNGEISGQMKQPLQANMPSISRGRTMIHHPGVANSSSTSPYSK 2248
QY 1269 K 1269
Db 2249 K 2249

Search completed: March 17, 2003, 12:27:34
Job time : 107.906 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:27:19 ; Search time 21.588 Seconds

(without alignments)
3552.777 Million cell updates/sec

Title: US-09-964-858-1

Perfect score: 8631

Sequence: 1 MNSYPSKLLPKDKHSLQLQ.....WVNLMLQQQQQQQQSSSQO 1664

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327	3.8	6281	10 US-09-815-242-12996	Sequence 12996, A
2	309.5	3.6	2437	10 US-09-815-242-5834	Sequence 5834, Ap
3	298	3.5	2025	10 US-09-815-242-5703	Sequence 5703, Ap
4	298	3.5	3158	10 US-09-815-242-12611	Sequence 12611, A
5	289.5	3.4	2344	10 US-09-815-242-12713	Sequence 12713, A
6	280	3.2	1139	9 US-09-820-843A-15	Sequence 15, Appl
7	280	3.2	2368	10 US-09-815-242-5635	Sequence 5635, Ap
8	280	3.2	2368	10 US-09-815-242-12389	Sequence 12389, A
9	280	3.2	3788	9 US-09-952-267-76	Sequence 76, Appl
10	276.5	3.2	1400	10 US-09-764-176-7	Sequence 7, Appl
11	271	3.1	1616	9 US-09-820-843A-16	Sequence 16, Appl
12	269	3.1	2843	9 US-09-987-482-1	Sequence 1, Appl
13	268.5	3.1	1863	9 US-10-022-819-2	Sequence 2, Appl
14	267.5	3.1	1863	9 US-09-734-672-4	Sequence 4, Appl
15	267.5	3.1	1863	9 US-09-982-828-6	Sequence 6, Appl
16	267	3.1	2434	10 US-09-815-242-5835	Sequence 5835, Ap
17	267	3.1	2478	10 US-09-815-242-5816	Sequence 5816, Ap
18	267	3.1	2478	10 US-09-815-242-12967	Sequence 12967, A
19	266.5	3.1	1111	10 US-09-815-242-12955	Sequence 12955, A

20	266.5	3.1	2843	8 US-08-681-219-32	Sequence 32, Appl
21	265.5	3.1	1863	9 US-09-734-672-2	Sequence 2, Appl
22	265.5	3.1	1863	9 US-09-734-672-6	Sequence 6, Appl
23	265.5	3.1	1863	9 US-09-982-828-2	Sequence 2, Appl
24	265.5	3.1	1863	9 US-09-982-828-2	Sequence 4, Appl
25	265.5	3.1	2285	10 US-09-932-183A-2	Sequence 2, Appl
26	264.5	3.1	2076	10 US-09-815-242-5815	Sequence 5815, Ap
27	264.5	3.1	2186	10 US-09-815-242-12913	Sequence 12913, A
28	262	3.0	665	9 US-09-820-843A-107	Sequence 107, App
29	262	3.0	5785	10 US-09-815-242-12610	Sequence 12610, A
30	258	3.0	1884	10 US-09-785-770A-17	Sequence 17, Appl
31	258	3.0	1907	10 US-09-785-770A-16	Sequence 16, Appl
32	257.5	3.0	1478	10 US-09-801-368-52	Sequence 52, Appl
33	253.5	2.9	2086	10 US-09-815-242-5639	Sequence 5639, Ap
34	253	2.9	2665	10 US-09-864-761-34248	Sequence 34248, A
35	253	2.9	3256	10 US-09-919-172-98	Sequence 98, Appl
36	246.5	2.9	1421	10 US-09-924-154-13	Sequence 13, Appl
37	245.5	2.8	1285	10 US-09-982-091A-2	Sequence 2, Appl
38	245.5	2.8	1805	9 US-09-820-843A-73	Sequence 73, Appl
39	241.5	2.8	1325	10 US-09-864-761-35612	Sequence 35612, A
40	240.5	2.8	1215	10 US-09-815-242-5908	Sequence 5908, Ap
41	240.5	2.8	1269	10 US-09-815-242-13113	Sequence 13113, A
42	239	2.8	2167	10 US-09-801-368-56	Sequence 56, Appl
43	238.5	2.8	1501	10 US-09-924-154-17	Sequence 17, Appl
44	236.5	2.7	1093	10 US-09-801-368-392	Sequence 392, App
45	236.5	2.7	1518	10 US-09-801-368-152	Sequence 152, App

ALIGNMENTS

RESULT 1
US-09-815-242-12996
Sequence 12996, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12996
LENGTH: 6281
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12996
Query Match 3.8%; Score 327; DB 10; Length 6281;
Best local similarity 17.7%; Pred. No. 1.2e-08;

Matches 296; Conservative 279; Mismatches 607; Indels 488; Gaps 71;

```

QY 42 NSPSPDLPPSSDITYTSEQ--DOEG-----KEEKD-----71
D 2459 NGSPVITNADYTAATQRYNNAETGLNGDNLATAKQAQKADLRQMTLHSDAQKQISITG 2518
QY 72 -----TAFOTSPDNFPLDMSID-IDOTIQHOQOPOOQOOLSTQDNKLDEFSFOT 122
D 2519 QIDSATQVTVQSVYKDATLQDNAMNOLRNSINKDKDVKASQPV--DADROKQNYNT 2575
QY 123 PMTST--IDLTKQNPVDKVNENHAPTYINTSPNKSIMKATPKASPKVAFTVPEI 179
D 2576 AVYNAEIIINATSGO-PLDPSATVQANQVST--NKALAGQNLAN-KQOETTANINOL 2631
QY 180 HHHPDRVEED-----OSOQKEDSVPELIIHQWD--PSQFNYS 218
D 2632 SHL--NNAQKQDINTQVTPNPNISTVNOVYTKAEQLDQAMERLINGIDRDQVKQSVNFT 2689
QY 219 DEDTNAVPTPEPLHTTKPFAO-LINKNNEVSEPE--ALTDKMLKRENSNLSLDE 273
D 2690 DAD-----PEKOTAYNNAVTAAENTINQANGTANOSQVEALSTVTTTKO--ALNGDR 2741
QY 274 KYNLYLPTNNNNKNVSDMSHLQNLQDASKNKTENIHLSPALKAPKNDIENPLSL 333
D 2742 KV-----TDKNNANQTLSTLD--NLNNAQKAVTGNL-----NQAHV 2778
QY 334 TNDISLRSSGSSQSLQSRNDNRVLESVPGSPKVNPELSLND-----GKGFSDVV 388
D 2779 AEVTAQAIQTAQELNTAMGNLKN-----SLNDKDTLTGSONFADA-- 2817
QY 389 ESLLPRDLSDKLETTKEHDAPEHNNENFDKSTJTN--KGOL-----LVSS 434
D 2818 -----DPEKKNAVNEAVHNAENILN-KSTGTNPKQDQVEAAMOVATKALNG 2865
QY 435 DQHLDSFDSYNTEDSILNLSASQS-----QISLMAEKQROTQEOBQTAABEE 489
D 2866 TQLEKAKGHANTAIDGLSHLTNAQKALKQLVQOSTVYEAQGNBQKANNVDAAMDCLR 2925
QY 490 TSPSDIKYKOEKPSKLEPKVYIKKEPVSAITKAPKEFSRILRIK--DEEIAEPAD 548
D 2926 QSLADNATTKO--NONTTDASQNK-----KAVYNAVTTAGGIIIDOTTSP- 2969
QY 549 IHPKKEENANSHVEDTDLKLLKALNDESDTTQNTSKSIRPHIDSQMLKEDSNDGDR 608
D 2970 LQPTVINQAGOVST-----KNALNGENLEAKQAQASQSL-----GSLDLNNAQKQ 3018
QY 609 D--NDDISREKSDIILNDVQSTSDIIGDKYGNSSSEITTYTL-----APPRSDN 655
D 3019 TVTDOINGAHATVDEANOIKONQONLNTAMGNLQALADKATKATVFTDADQAKQOAVN 3078
QY 656 NDKENSKSLEDDAN-----NESLQQLVPHPKKEDDSILANS--- 693
D 3079 TAVTNAENIISKANGNATQAEVEQAIKQVNAKQALNNGANQAHKDDATLALINSNDL 3138
QY 694 NIAPPELTLPVVEANDVSSFNVDYTKF---DAVSSFEESLSREHETDSKPINFISIMH 749
D 3139 NQAOAKDALQOQVQNTTAVGVNVKQTAQELNNAMTQLKGIAIDKQGT----- 3186
QY 750 KQEKQKHHQIHVPTKQIIIASTQYQKNEQESRYTSPKVIIPNAIQKFKREVNVMSRRVY 809
D 3187 ---KADGNFVNADPPKQNVYNOAVAKAEALISATPVDVVTPEI-----TALNRY 3234
QY 810 SPDMDDLANSQFLPELSEDSGFKDLNFANYSNNTNPRSFPLST--KNVLSMIDNPRV 867
D 3235 TQAKNDLN-----GNTN-----LATKQVNOHAIDQLPML 3264
QY 868 VEPPEKSYAEIRNARLSANKAAPNOAPPLPPOQOPSTRSNSKRVSRFRVPEIR 927
D 3265 NQAOQDEISKQITQO-----TLVFNVAIQQAATTLNDAMTQLQGIANKAQITG 3314
QY 928 TSSALAPCDMYNDI-FDEGAGSKPTIKAGM---KTLPSMDKDVKRIILNAKKGVTODE 983
D 3315 SEN-----YHADDTDKQTAVDNAVTAKAEELLKQTTNTTMDPNTIQAQLTVVNDINQAL 3367

```

```

QY 984 YINAKLVQK-----PKNSIVTDPREYEELOQTASINATIDSSITYGR 1028
D 3368 NGNQKLADAKQDAKKTTLGTLIDNLNDQAKQALTT-----QVQADRIATVNVKQNAQVL-- 3421
QY 1029 PDSISTMDLPLSLDELKPPRTALLSADRLFMEDQVNRPLRSNVLYHFGAATNSSMLPE 1088
D 3422 -----NNAMTNLNNALLODKTETLNSINFETDADAKKDAVYTNAAVSHAEGILSKANGSNASQ 3476
QY 1089 PDFE-----LINSFARNVSNNSDNVAISGNAS-TISFNQDLNMFDDQATTQOKTOEOPASK 1143
D 3477 TEVEQAMORVNE-AKQALNGNDVQORAKKDAKOVITNANLNOAQKALQOQVDAQVTA 3535
QY 1144 SANTVRGDDGLASAPETPTPK-KESISSKPAKLSSASPRKSPIKISPVHRIKNGS 1202
D 3536 NVMTIK-----QTAQDLNQAMTQLKOGIADKQO-----TANGN 3569
QY 1203 IAGIEPLPKATIKPKKSFQSGNEISNKKVDRGDISPSSGSEHQNDPMSVPSQYTDATS 1262
D 3570 FVNAD-----TDKQAVNNNAVA-----HAQOI 3593
QY 1263 TVPDEKNDQVKKPREKQK--QKHNNHHNNHHNKKOTDIPGVNDELPRDVLQERKLFER 1320
D 3594 GTPNANVDRQVQAQALQOVQAQKGLNGNHNLOVAKDMANTALDQLEP----- 3640
QY 1321 VLGIKNINLPDINTHGRT-LTLDNGVNCVTPPEYNMDDHVAIGKEFELTVADSLFEI 1379
D 3641 -----MLNPOKRTALKDQVSHAELVGVNAI--KQADALNNAMG----- 3678
QY 1380 LTLK-----ASYEKPRGTVTEKKAUVKSR-----NRLSRLEFGSDIITTTKVFVTEVK 1429
D 3679 -TLKQOIQANSQVPOG--VDETQADODKQOQAVYNNANQAOIANG-----IPTPV- 3725
QY 1430 DTWANKFAPDGSFARCYIDLOQFEDQITG--KASQFDLNCFNEMETMSNGNPKRGKPY 1487
D 3726 -----LTPD-TVTAQVNTMNAQKDALNGDEKLAQAKQKELANLDLTRLQDQORALNR 3778
QY 1488 KIQOLEVKMLYRSPRELPTPSIRSAYESTIENLNEONNFEGLYHOE 1537
D 3779 QINQAQAL-----ATVEOTKQONAMQNVNTAMSNLKGIAIANKD 3814

```

RESULT 2

US-09-815-242-5834

Sequence 5834, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA-011A

CURRENT FILING DATE: 2001-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-05-23

PRIOR FILING DATE: 2000-05-23

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-10-23

PRIOR FILING DATE: 2000-10-23

PRIOR FILING DATE: 2000-11-27

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR APPLICATION NUMBER: 60/269,308

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5703
LENGTH: 2025
TYPE: PR1
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(2025)
OTHER INFORMATION: Xaa - Any Amino Acid
US-09-815-242-5703

Query Match 3.5%; Score 298; DB 10; Length 2025;
Best Local Similarity 18.9%; Pred. No. 9.4e-08;
Matches 333; Conservative 246; Mismatches 663; Indels 524; Gaps 79;

17 LQLOPOSSASIFNSPKRP-----LNPRRTSKSLDP-----NSSSD 54
18 IODQOQTEAGSKFNEKRDQAYQAAVQAHAKDLNQTNPDLKAYEQLTQGVNQAKD 447
55 TYTSEODEKEKEDKDTAFQTSFDRN-----FDLNNST---DIOQTIGHQOQRP-- 102
448 NLHGQKADKQAHAVYDNLQNLNSLNFQRALESQIINNAATRDVEAKKLEAQAQADAM 507
103 -----QQOQLSQTDNNLIDE-----FSQTPMTSTLDLTKQ--NPTVDRV----- 140
508 QALRNSIQDQOQTESGSKFNEKRDQAYQAAVQAHAKDLNQTNPDLKAYEQLTQGVNQAKD 567
141 -----NENAPTYINTSPKSKIMKATPKASPKKVAFTVTNPELHHYRPD 184
568 VTTAKDNLHGQDKLARDOQAQVTTYNALRN---LNHAQOQALTDALINAPTRTEVAQHVC 624
185 NRVEE---DOSQOKEDSEVERPLIOHOKWDSOFNYSDEEDINASVPRPRPLHT---TKP 237
625 TATELDNAMELTKKLVQDN-----TKAQRPYTEASTDKKEAVDAQALQAEISTDP 676
238 TFAQLKNNEVNSEPALTKMLKRENFSMLIDEKYNLYLSPNNN-----NSKN 289
677 TNGSNANK---DAVEQALT-----KIQEKVN-----ELNGEYVAEAKQAQAKQ 716
290 VSDMDSHL-----ONLQDASK-----NKTENINHLSPALAPKN----- 324
717 TIDDLAHLNMAQIDATKQINQIDATKLPRIAEVLVQATQOLNOSMDLOAQVAEHNAVEQTV 776
325 -----DIENPL-NSLTNADISLRSSGSSOSLSQSLRN-----DNRVLESYGS 366
777 DYTQADSDKQAAVQKQALAEANVLKQNSKKQVDDALQNLINAKQALNGDEVALAKTNG 836
367 PKKVPGLSLN-----DGIKGFSDEVESLRLPRLDSRDLKLETTKEHDAPEHNNEFID-AK 421
837 KHDIDQALNNAQDQDKGRIDQ-----SHDL--NQIQQIVDQAK 875
422 STNTNKGQLLVSPDHLDSFDRSYNTHQESILNLSNASQ--SQISLNALEKOROTQEOEQ 480

876 ALNRAMDQ-----SQETISNEGRTKST-----NYVNADTQKQVYDAVADKAKQALDKST 927
481 TQAAPEEETSFSDNIKKQEKSKLSEFVKYTIKKEPQSAE-IAKAPRESSRLIRKIN 539
928 GONLTAEQVIRKID-----AVTAAKKALNGEERLNNRRSEALQRI----- 967
540 EDEIAEPADIIHPKKEANEANSHVEDTALLK--KALNDEESDPTQNSKMSIRFHIDSDW 597
968 -DQLTH---LNNAQROLAIQDINNAETLTKASRAINRATKIDNMG-----AVQOITIDGH 1019
598 KLESDNGDREDNDDISPEKSDIINDVSQTSIDIIGDKYXGSSSEITTKTLAPPSDND 657
1020 LGVISSTVYINADDNLKANYNAINAAHELDKVGQNAIAKAEABQLQNIITDAQNALNG 1079
658 KENSKSLDDPAN-----NESIQOOLEVPHRKEDDSILANSNTIAPPELTPVYEADN- 710
1080 DONLANAKDKANAEVNSLNGLNQOQODLAAH-----NAINMADTVSDVTYDINNQIIDLDA 1134
711 -----YSEFNDVTKT--FDAYSSEESISREHETDSKPIINFISTWH 749
1135 METLKLHVDNEIPNAEQVTYVQNMADNNAKTINFDQAKRLANTLLNSDNTNVMDIN----- 1188
750 KOEKOKKHQIHKVPYKQIITASTYQYKNDQESRVTSDK-VKIPNAIQKFKFEVNVMSRRV 808
1189 -----GAIQAVKDAIQNLNGEORLQEAQKDAIQVNNKVLADKLEKEIASN--- 1233
809 VSPDMDDLNVSGFLEPSEDSGFKDLNFANYSNNNRRSFRFTPLSTKVNLSIOMDPVNV 868
1234 -ATDQDKLIAKKAEELAN-----SIININNKATSNODYSOVQTAGN----- 1274
869 EPPEPKSYAEIRNARRLSANKAAPQAPPLRPQ--RQPSSTRSNKSKRVSRFRVPTF--- 923
1275 QAIQEVHANEIPKA-KIDANDVDKQVQALIDEIGRPNLIT--DKQKQALDRINOIQQ 1331
924 -----EIRTSALAPCMYNDIEDDFGA--GSKPTIA----- 955
1332 GHNDINNALKREAIQAKERLA--QALODIDLVYAKEDAKKIKALANAKRQDQINSND 1389
956 -----EGMKTLPMSMDKDYKRLINAKKGVTODEY-----INAKL-----VDQKPKNS 998
1390 LTPBEKAKALKEIDEAEKRALQNVENNAOTIDQNLNGILGLDIDNTHTWEYEDDQPAVNE 1449
999 IYTDPEDRYEELOQTASIHNAITDSSISYGRPSIST--DMLEPYL-----SDELKRPPT 1049
1450 I-----SEATPEQLLVNCELIVHHRDITTEQDVLAHINLIDQLAEVIDPST 1497
1050 ALLSADRLEFMEQVHPLNSNSVLVH--PGAGATNSMSMLPEPDFELINSPAR--NVSNNS 1105
1498 ATIS-DSLTAKVEVTLIDGSKVIVPVKVEKELSVVQQAIESIENAVQOKINEINS 1556
1106 DNVAISGNASTIS-FNOL-----DNFDDQATIGOKIQOPASKSA- 1145
1557 VTLTIEQKEAALAEVKNKQQAIDHINNAAPDVHSAVEELIQOQEAHIEQFPNEQFTIEQAK 1616
1146 -NTVAGDDGLASAPETRPPT-----KKSISSKPAKLSASAPR----- 1184
1617 SNAKSIDDAIQHMDIEKARTDLTFKKEQEA--AKLNQKQALQIOLQRAQSIDEI 1672
1185 -----KSPIKIGSP-----VRYIK-----KNSIAGIEPIPKATKPKKSF 1220
1673 TEOLRFQAKQKAAANPTAKELAKRQKQEAISKIDPESNEKMSIRNS-IGRADEQQAAMN 1731
1221 QGNELSNHKVRD-----GGI-----SPSSGSEHQHNS 1249
1732 QINELVLETTIRIDINNAHPLQOQVEAALNNGIARISAVQIYVTSDRAKOSSSTIGESHSH--- 1788
1250 MVSVPQYTDATSTVDEKNDVQNHKPREKOKNHNHNHNHNKOKTPIPVVDEIDPVD 1309
1789 ---LTIGV--GTAHPFNSSTIGHKKILDEDDIDPCLMHRP---SNFGVNIKALICV 1840
1310 GLOERKGLFFRYLGIK-----NINLPDINTHKGRTLLDNGVH--CVTTPPEYAMD 1358

QY 983 EXINAKLVDOCKPKNSIVTPEDRREELQOTASINHAATIDSSI-----YGRPDS 1031
DB 1953 DSISSISVSDSSSTSTSLSDSMQSOISTASAGSLSTSLSDMSASAGLTQSIS 2012
QY 1032 ISTDMIPYLDELKPPYLLSADRLPMEQEV---HPLNSVYLHPGGAATNSSMLPE 1088
DB 2013 VSTSL--STSDSISTSTISISOSQAVESSTSDSTSLSDSLSSTSGSTSSSTSTST 2070
QY 1089 POFELINPARNV-----NSDNVALSGNASTISFNQIDMNDQATIG 1133
DB 2071 EBLSTNSGSTSVSESLSTSGSGSTSVSDSSSTSSLSSTGSTSVSDSTSMSESNASIS 2130
QY 1134 OKIOEPASKANTYVAGDDGLASAPETPTP--TKESISSKPAKLSSAPKPSIKIG 1191
DB 2131 MS-QSISGSGSTISSTSESLSMGSGTINSTSVSDSDSISTSGSMSTIHFTSLST 2189
QY 1192 SEVRYIKKNGSIAGLEPIPKATHKPKKSPQGNISNKKYVDGDISPS---SGSEHOHNP 1248
DB 2190 SGLMSLSDNSMSDSDSVSISASEMSASMDSVMSDSTSSMSMSKSTSESNISHP 2249
QY 1249 SMVSVPOVTDATSTVPREDKVOHKKPREKOKHHHHHHHHKQTDIPGVVDEIDP 1308
DB 2250 SN-SMSTQHTSTSTISSTESIAPTNESQSTLSATSVSKHDAE---PAOSEERLPD 2304
QY 1309 VG--LOERKGL 1317
DB 2305 TGDSTIKONGL 2315

RESULT 6

US-09-820-843A-15
; Sequence 15, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: M. genitalium
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: cycladherence-accessory protein
; NAME/KEY: misc_feature
; OTHER INFORMATION: g111046012
US-09-820-843A-15

Query Match 3.2%; Score 280; DB 9; Length 1139;
Best Local Similarity 19.0%; Pred. No. 4e-07;
Matches 225; Conservative 188; Mismatches 435; Indels 334; Gaps 53;

QY 39 PRTNSKPSLDPPSSSDTYTSEDOQEKKEKKTAFQTSDFRFDLNSI--DIQOTIGH 96
DB 184 PEVKQESVVDQSSDDYFAKQPT-----DENYGFNDNDLPPVKKOP-ES 226
QY 97 OOOQPOOOQOOLQOTNNLIDEFSGFQTPMTS--TLDLTQKNTVD---KVNENHAPTYINTS 152
DB 227 VVDQPSDDHFAKQPESTSDSYSFDSDLQPTLD-----QPSLDHNYQYNVDH----- 274
QY 153 PKSTINKKATPKASPKKVAFTV-----TNPETIHNYPDNRVEEDOSQOKE--DSVEPPL 204
DB 275 --HEELKPVAEEOHNYQVGFQVQANLIDNNEEIQPTAEKKVTTDFESKQAVVDYSOLPI 332
QY 205 IOHOMKDPQOFNYSDDEDTNAVPTPPLTTKTFQQLNKNNEVSE--PEALTMKML- 261
DB 333 DTDO-QDQTFSSSFE-----TQPTVEQFDQVNSEVNDQEPFLTKPEVLE 377

QY 262 ----KRENFSLDEKVNLYLSPFNNNNSKAVSDMDSHLQNDQASKNKT-----N 309
DB 378 SFTNKQDVYETSDLNSESNIY--SENNKQATNNDSLNEFTQLNNSNETASDYHYEKS 435
QY 310 ENIHNLSPALKAPKNDIENPLNSLTJNADISLRSSGSSQSLQSLRNDNRYLESVPGSPKK 369
DB 436 EPIHDYKF-----GSDLSQS-----NSNNLES---EPVK 462
QY 370 VNPGLSLNDIKGFS---DEVESLPRDLSRKLKLETKHND---APEHNNEFIIDAKS 422
DB 463 FNSETAPDAHFESQSEPVQDYOVDIYQNELKPTLDQPSDDYFAKQPTDENGFQNDLP 522
QY 423 TNNKCOLYV---SSDHL-----DSFGRSYNHEQSTILNLSNASQSLANLEKORQT 475
DB 523 PEVKQESVVDQSSDDHFAKQPESTSDSYSFDS-----LPQPTLDQPSLDHNYQYNFD 577
QY 476 QEOQOTQAAPEBETSEFSNRIKVKQEPKSLFEVKTIKKEPVASATEIKAPKEFSRIL 535
DB 578 HHEELKPVV--EEOHNYQVGFQVQANLIDNNEEIQPTAEKEVTT-----DFESKQA 626
QY 536 RIKNEDEIAPADIHKKEKMEANSHEDTALLKALNDEESDITQNSTKMSIRHIDS 595
DB 627 QVVDYQVL--PIQTDQDQDTTFSSSFEPTQPTV-----EOPDVQVSEVNDQFKPEITK 676
QY 596 DMKLEDSNDGDRDNDDISRFEKSDILNDVQSTDIIG--DKYNSSEIYTKTILAP----- 650
DB 677 EPLVLESS-----FNKQDVYETSNYNNLQKPIQSDNKTITTTKSSPOIPT 723
QY 651 --PRSDNDEKSKSLEDPA--NNESTLQQLLEVPHRKEDDSILANSSNIAPPELTLPV 706
DB 724 TLPISFVSNRIEYKPVETLALDKKESQEQITINSTIATEDSKTLAKT-----LSVQL 775
QY 707 EANDYSFNDVYKTPFAYISFEESLSREHETDSKPINFISIMHKQEKHQHVKVPTQ 766
DB 776 QINLSNOSTIVT-----SESVRLDKRKDDLTITFV---NSEDOQPIEVF-VYAKE 822
QY 767 IIASYOQYKNEQESRYTSQVKIPNAIOFKK---FKEVWMSRRVSPMDDLNYSQFL 822
DB 823 PVEHSTIQNKOS---VEDSELDN---FNKKDLYIIEELRGELNPIINFDAIFQMN 876
QY 823 P-ELSEDSGFKDLN--FANYSNNTNRPRTSFTPLSTKNVLSINDNPVVEPPEKSYAET 879
DB 877 DYQMSVQSFIIHLNDEVTNKNQISE----- 902
QY 880 RNARLSANKAARNQAPPLPQORQPSSTRSNSKRVSRFPVFEIRTSALAPCDMYN 939
DB 903 ---RYLLIKKELOSELRLIDQENLNVQFNNAKNL-----TTLQKEEMIR 945
QY 940 DIFDDFGAGSKPTIKAGMKTLPMSMDKDVKRLINAKKGYTODEYINAKLVDOCKPKNSI 999
DB 946 SLASDFALAKPNSNYEQLO-----KSGELMR--HYQRAITENE-----KR--- 984
QY 1000 VDPEDRYEELQOTASTIHNAATIDSSIIYGRPDISTDMLPYLDELKPPYLLSADRLFM 1059
DB 985 ---IESIOGSLKQDKLKTYYNSCCE-----IMNNIKLIDNTL-----RFA 1020
QY 1060 EOEVHPLR-----SNSVYLHPGAGAAATSSMLPEPDEFILNSPANNVNSNDNVAISG 1112
DB 1021 KKEKDPILLSNFDSDVTDNGLVEP-----NQLMDLIDL-----SNTFDNIS-NE 1063
QY 1113 NASTISFNQIDMNF-----DQATIGQIKQEPASKSAN 1146
DB 1064 QLDQFTIENMDRNIDFEFEGFNNDPVIDAKVMDSKAFSVN 1105

RESULT 7

US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: fastseq for windows version 4.0
SEQ ID NO 5635
LENGTH: 2368
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match 3.2%, Score 280; DB 10; Length 2368;
Best Local Similarity 18.3%; Pred. No. 1e-06;
Matches 277; Conservative 242; Mismatches 574; Indels 418; Gaps 61;

QY 64 KGEKKKPT-----AFQTSFD-----RNFDDLNSIDI-QOTIOHQOQOQOQOQOQO 109
DB 885 KSKQVADPTKSKVLDKINAIQOAKVPADVEVENAVTRKQEOIYNSASTTEKQAAV 944
QY 110 TDNNLIDFESFQTPMTSLDLTKQNPYDKVEN--HAPTYINTSPNKSIMKA--TPKA 165
DB 945 TE---LD--TKQGEARTMLDANNTSDVTTAKDNGIAINQOAAITTKSKAKELAQKA 999
166 SPKKAFTVNTPEIHVHPDNRVEEDQSOQKEDSVPEPLIQHQKDPQSOFNYSDED---- 221
DB 1000 SERKAIAIAMN-----DSTTEEQ--QAAKDKVDAVVTANADIDNAANTVDNNAKT 1049
QY 222 TN-ASVPTPLHTTKPTFAQ-----LTKNNEVNESEPALDMMKUREN----- 265
DB 1050 TNEATIAITPDANVKTAKQAIADKVOAQETAIDANNGATTEERAAAKQOQVTEKTTAD 1109
QY 266 -----FNLSDIDEKVNLYL-----SPTNNNSK-----NVSDMDSHLONLQDASKN 306
DB 1110 TAIQGAHNTAEVEAKNAEIKAEIAIQPATTTKQAKAQAIAIKAEKRTAIAOTODITAE 1169
QY 307 K---TNEINHLSEFALKAPKNDIENPLNSLTNADISLRSSGSSQSSLOSRLRDNVLESV 363
DB 1170 EIAANANAVDN---AVTQANNRIE--AANSQNDVD---OAKTGEASIDQV-----T 1213
QY 364 PGSPKAVPGLSLNDGICIGFSDE--VESILPRDLSRDKLETTKEADAEHNENFIDAKS 422
DB 1214 PTVNKATAVTDAKNNTIAATDNGVDV--AKDAGKNSIQSPATVAVSNKKNVDQAV 1271
QY 423 TTNNGQLVSSDDHLDSFDRSVNHTEOSILNLSASOSQISLALAEKOROTQOEOQTO 482
DB 1272 TTQN-----QAIDVTTGATTEKNA--KDLVLAKAEKAYDILMAOTT 1313
QY 483 AAPEEETSFSDNIRKVKQEPSNLEFVKVITIKKEPVSAETIKAPREFSSRLIRIKNEDE 542
DB 1314 -----NDVNTQIKQOAVADVQGITADTTIKIDVAKDELATKARE-----QKAL 1354
QY 543 IAEPADIPKKEKNEANSVHEOTDALLKALANDEESDTTQ--NSTKMSIRFIHDSWKLE 600

DB 1355 IAOADATTEEREQANQOV---DAELTOGNONIEAASIDVDVNTAKDNAIQADIPQAST 1411
QY 601 DSDNDGREDNDNISREKESDILNDVSQSDIIGDKYG-----NSSEITTKILA 649
DB 1412 DVKTNAARELLTEKONKITEIILNNETTNEEKNGIGVRAAYEEGLNINATTTGDTT 1471
QY 650 PPSDNNKE---NSKSLDEPANNEILOQOLEVPPTKEKDSILANSSNIAPEEETLPV 706
DB 1472 TAKDVAQKVQGLHANPVPKPKAGKTALQDAADKKTQTEGTFNNAQOEL----- 1520
QY 707 EANDYSEFNDVTKTDVASSFEESLSREHEDSKIPINFIISTHKKOEKQKHQVPT-- 764
DB 1521 -----NDAKQEVDTTELQAKTINIDQSSDEYVDNAV-----KKGAKAINAVKTF 1565
QY 765 -----KOILSYQOQYKNEQESRVTSQKVPINAIQO-----FKKREYVMSR----- 807
DB 1566 EYKKDALAKTEIAYNAKATEADNSNASTSSEIAEAKQKLAELKOTADONVQATSKDIE 1625
QY 808 -VVSFMDLNVSOFLPELSEDSGFKDLNFA-----NYSNNTNRPSPFTPLSTKNYLS 859
DB 1626 VQIHNDLDNIN-DYITLPQKKESATTDL-YAADQKKNNISADTNATQDEKQOAIKQYDQ 1683
QY 860 NIDNDPNVVEPEPKSYAEIRNA-----RRLSNKAAPN---QAPPLPQOPSSSTR 908
DB 1684 NVQT-----ALSSINGVNDGVDALTOGKAIDAIOVDATYVKPRANQVIDAK 1732
QY 909 SNKRVSRFRVPTFEIRTSALAPCDMYND-----IFDDGAGSKPPIKKEGKTLP 963
DB 1733 AETKE-SIDQSDQLTAEREKTEALMIAQITDOAQGITDATTAEVAKARQGLEARDN 1791
QY 964 MDKDDVR-----ILAKKGV-----TQD----- 982
DB 1792 IQIDSTEROKAIELETAALDQIEAGVNDADATTEEKFAFTALDILSKATEDISDQTT 1851
QY 983 -----EYINAKLYDQKPKKNSIYVDPEDRYEEL--QQTASHNATIDSIYGR 1028
DB 1852 NAEIATVKNSALEQKAKQRTINPVKKNL-----BAIREVVKQIETIKMAADAS---A 1903
QY 1029 PDSISTMLPY---LSDELKRPPT-----ALLSADRLFMQOEYVPLRSNSVLVHP 1075
DB 1904 KEIARTDLGRVDRPADKLDKQNTQTEVAELQNTVIRPAIEAVPQNDNANDNTNGSDNN 1963
QY 1076 GAGAAATNSMLPEPDELINSPARVNSNNSDNVAISGNASTISFNQOLDMNFDDQATTQOK 1135
DB 1964 DATANSNANATPE-----NTGQPVWTESTDN-ANADTSSTTTNQ-----NDAAT 2007
QY 1136 IQQPAKSAANTVRGDD-----DGLASAPETPRTPTKKESSKPP----- 1175
DB 2008 -GETTATISANSATDNDKPKQANNNSSADTSTNSPTMDNDVTSKPEVESTNNGTQDP 2065
QY 1176 -AKLSSASPRKSPKIKIGSPVRVIRKNG----- 1201
DB 2066 ATEADNATPRAEATNTNNSTTTATNTENAPTESTATAPPTASTGAESSADSKONASVNDKQ 2125
QY 1202 -----SIAGIEPIKATHPKPKSFQGN-----ISN 1227
DB 2126 NAEVNNASAESQSTNGKVAQPKSENKAKAEKDGDRDSTNGSWESTETELPSADITEPKVSS 2185
QY 1228 HKVRDGGISPSGSEHQOHNPSMVSPQYTDATSTVDENKQDQHKRREKQKQKHHNRH 1287
DB 2186 NTSKDEESTISQTAEOHN-SDTVNVAENAKDSGQNDV--TDVSNKPSSTKSPSEAKDKA 2242
QY 1288 HHHHHKOKTDI 1298
DB 2243 TSTEDQKADM 2253

RESULT 8
US-09-815-242-12389
Sequence 12389, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12389
LENGTH: 2368
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match 3.2%; Score 280; DB 10; Length 2368;
Best Local Similarity 18.3%; Pred. No. 1e-06;
Matches 277; Conservative 242; Mismatches 574; Indels 418; Gaps 61;

64 KKEKKKDT-----NFOTSFDRNFDLNSIDI--QOTIGHOOQPOOQOOLSQ 109
DB 885 KSKOEADPKSVYLDKINIQOAKVPADAEVENAYNRKOEIONSNASTEEKQAV 944
QY 110 TNNULDEFSPMTSTDLTKONTYVKNEN--HAPTYINTSPKSIKKA--TPKA 165
DB 945 TE---LD--TKOEAETNLDANTNSDYTAKNONGIAINQOAAATTKSDAKAELAQRA 999
QY 166 SPKVAFTVNPENIHYPDNRYEEDOSQOKEDSEVEPPLIOHQMKDPQSFNSYSD 221
DB 1000 SERKTAIEAMN-----DSTEEQ---QAKDKVDQAVVTANADIDNMAANDVDNKT 1049
QY 222 TN--ASVPRPPLHTTKPTTAQ-----LTKNNEVNSPELALTMKLRN----- 265
DB 1050 TNEATIAITPAANVPKATQKALADKVQAOETAIDANNATTEEKAAARQOYVTEKTTAD 1109
QY 266 -----FSNLSDSEKYNLTV-----SPTNNNSK-----NVSMDSHLOMIDASKN 306
DB 1110 TAIDGAHTAAVEAKKNAIEIAKIEAIOPTATTKDNKAKAIAITANKRTAIAOTODITAE 1169
QY 307 K---TNEINHLSPALKAPKNDIENPLNSLTJNADISLRSSGSSQSLQSLRDNRYLESV 363
DB 1170 EIAAANANVDN---AVTQANNIE--AANSQNDVD---QAKTTGEASIDQV-----T 1213
QY 364 PSSPKVANGSLNGCIKFSDE--VESILLRPLSRDKLETTKEHAPENNNENFIDAKS 422
DB 1214 PTVNKAATVADKAKNITATADNGVDV---AKDAGKNSIOSITOPATVAVSNANNDVDQAV 1271
QY 423 TTNKGQLLVSSDDHLDFDSRYNHTEOSILNLSASOSQISLNLKOROTQOEOQR 482
DB 1272 TTON-----QALDNTTGATTEKMAA--KDLVLKAKKAYQOILNAQOTT 1313
QY 483 AAEPEETSFSDNIKVKQEPKSNLEFVKVYTIKEPEVSATEIKAKPREFSRILIRKNEDE 542

DB 1314 -----NDVTQIKDQAVADVQGITADTTIKDVAKDELATKARE-----QKAL 1354
QY 543 IAEPAIDHPKKEANSVHEDTALLKALNDDEESTTQ--ASTKSIHFHDSMKLE 600
DB 1355 IAOATAATTEKEQANQV---DAELTQGNQNIENAGSIDVNTAKNAIAQALDIPQAST 1411
QY 601 DSDMDREDNDDISREKSDILNDVSGTSDILGDKRY-----NSSSETTTTLA 649
DB 1412 DVNTNRAELLETTEKQNTITLNNNETTNEEKNDIGPVAAYEEGLNNINATTTDDVT 1471
QY 650 PPSDNDKE---NSKLEDPANNESLQOOLEVPHTRKEDDSILANSSNIAPEELPLPV 706
DB 1472 TADTAVQKVOQLHANEVKKRPACKTALDQAADKKTQIEQTPNASQOEI----- 1520
QY 707 EANDYSSFNVTTPFAYSSFEESLSHEHTDSKPIFISIMHKOEKKHQIHKVPT-- 764
DB 1521 -----NDAKQEVDELQAKTNTIDOSTDEYVDNAV-----KEGAKINAVKTF 1565
QY 765 -----KQTIASVOQKNEQESRVTSDKVFIPNAIQ-----FKKKEVNVMSRR----- 807
DB 1566 EYKKDALAKIEAAYNAKVTLEADNSMASTSSEIAEKAKLAEIKOTAPQNVNQTASKDIE 1625
QY 808 -VSPMDDLNVSOFLPELSEDSGFKDLNFA-----NYSNNTNRRSFTPLSTKNVLS 839
DB 1626 VOIHNLDLNNIN--DYTIPTGKESATIDL--YAYADQKKNNISADTNATODEKQAIKQVDQ 1683
QY 860 NINDNPVVEPPEKSYAEIRNA-----RLSANKAAN---QAPPLPQORQPESTR 908
DB 1684 NVOT-----ALESINNGVDNGVDALTOGKAALDIOVDTVKPKANQVIDAK 1732
QY 909 SNGNKRVSRFPVPTFEIRRTSSALAPCDMYND-----IFDDFGAGSKPTIKAEKMTLPS 963
DB 1733 AEETKE--SIDQSOQLTAEKTEALAMIKOITDOAKQITPATATTAEVKKAKQGLEAFDN 1791
QY 964 MDKDYKR-----ILNAKKGV-----TOD----- 982
DB 1792 IQIDSTEKQAIETALDQIEAGVNDADATTEKEAFTNALIEDILSKATEDISDQT 1851
QY 983 -----EYINAKLVQDKPKKNSIYTDPEDRKEEL--QOTASINMATIDSSITYGR 1028
DB 1852 NAEIATVKNALQKLAQRINPVYKKNAL-----EAIREVANNQIETIKNADADAS---A 1903
QY 1029 PDSISTDMLPY---LSDELKKPPT-----ALLSADRLFMEQEHPLRSNVLVHP 1075
DB 1904 KEIARLDLGRFPDRFADKLDKQITQNTFEVALQWNTTIAIEIAYQNDPNPANDNTSSGSDNN 1963
QY 1076 GACAATNSSMLPEPDELINSPPARVNSNDNAISGNASTISFNOJDMNFDQATTGOK 1135
DB 1964 DATANSNANATPE-----NTGQPNVTESTDN--ANADTSSSTTTNNQ-----NDAAT 2007
QY 1136 IOEQPASKSANVYRGD-----DGLASAPETPRTTKRESISSK----- 1175
DB 2008 --GETTATSANSATDADANDKPOANNNSADTSTNSPTMNDVTSKPEVESTNGTTDKP 2065
QY 1176 -AKLSASPKRSPPIKIGSPRVIRKNG----- 1201
DB 2066 ATEADANATPAESATNNSTTTATNENAPTESTATAPTTASTGAESSADSKDNASVNDKQ 2125
QY 1202 -----STAGLEPIPKATHKPKKSFQNE-----ISN 1227
DB 2126 NAEVNSAESOSTNGKVAQKSEKKAKEKDRGSTRSQSWESTTETLLPSADITPEKVS 2185
QY 1228 HKVRDGLSPSSSEHQHNPVAVSPQYTDATSTYPPDENKVOYHPRKQOKHNNRH 1287
DB 2186 NTSKDKKEESTTSQTDAEQHN--SDTNVANSNADKSEGNVD--TQVSNKPTSKPSEAKDKA 2242
QY 1288 HHHHHKQXTDI 1298
DB 2243 TSTEDSQKADM 2253
RESULT 9
US-09-952-267-76


```

QY 101 P00000LSDTNNLIDESFOTPMSTLDTLRONPTVDKVNENHA-PTYINTSPKMSIMK 159
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1043 PS0NERMARPKPHIIEDEL-----K050R0SRM0STITYPYTTESTD0KHL-- 1087
QY 160 KATPKASPKKVAFTVNTBEIHHPDNRY-----EED0S0OKEDSVE---PPLI0H0M 209
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1088 KFOPIFG0Q0ECVSPYRSGANGSETRNGVNGINGVNSQSLCQ0DDYEDDKPTVYSEK 1147
QY 210 KDP0GFNTSDEDTNASTVPTPLHTTKPTFAOLNKNNEVASEPALDMDKIKRENFSL 269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1148 SEED0HEERPTVNSI-----KYNEKRNHVQ0PIDYSLKVA--TDI 1187
QY 270 SLDEKVNLYLPTNNNNKNSVSDMDSLONQDASKNKTNN--IHLSPALKA--PKNDI 326
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1188 PSS0K0SFSSKSSG0SKTEHMSSEENSTPSSNAKRONQLPSSA0SHSG0P0KAA 1247
QY 327 ENPLNSLTNADISLR-----SSGSSQSL0SLRND--NRVLESVP0GSPKRVN0L 374
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1248 TCKVSSINQETIQYVCVEDTPICFRCSLSSLSAEDIGCQNTQ0EADSANTLQIAEI 1307
QY 375 SLN0GIK0FSDEVE---SLPRDLSPDKLETTKEH0APEHNNENFID-ANSTVNNKQ 429
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1308 KEKIGTRSAEDPVSEVPASQHPRTKSSRL0SSLSSESARKAVEFSSGAKSPKSGAO 1367
QY 430 LLVSSDDH-----LDSEFDRSYNHTEOSILNLSASQSO-----ISLN 467
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1368 TPKSPREHYVETPLMFSRCHSVSSLDSE-----SRSIASSV0SP0CSGWSGSIISPS 1421
QY 468 AL-EK0R0T0E0T0A0APEETSFS0NIKVK0PKSNLEFVKYTIKPEVSAATEIKAP 526
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1422 DLPS0G0T0P0PSRCKT0PP0P0TA-----QTKREVPKN-----KAPTAERESGP 1467
QY 527 KR-EFSSILIK-----NEDEIAPADHHPKE 554
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1468 KOAAVNAV0V0YL0PADTLLHFT0ESTPDGFCSSSL0SLSD0EPF0K0VEELRIMPP 1527
QY 555 NEANSHVEDALALKALNDEESDTONSTKMSIRFHID0MKLEDSNDGREDNDIS 614
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1528 VQ0ENDNGMETSE0PKESNEN0KEA0KT-----IDSEK0LLD0S-----DD0IE 1573
QY 615 RPEKSDILNDV0S0TDIIGDKYGNSS0ITTKTAPRPSDNNK0ENK0SLEDPANNE0I0 674
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1574 ILEEC-IISAMPKSSRAKRAKPA0TASKL-----PPVARKPS0LPVYKLLP0QNR-LQ 1625
QY 675 Q0LEVPHTKEDD-----SILANSSNIAPPEEL----- 701
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1626 P0KHVSFTPGDDMPRVYVEGTPINFSTATSLSLDTLIESPPNELAAGEVGRGA0SGEPE 1685
QY 702 ---TLPV0EANDYSFNDVTKTFDAYSSFEESLSREHEDSKPINFISIMH0K0K0KHQ 758
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1686 KRDTIPTG0RSTDE0AG0KTSV-TIPELDONKAE0GLIACIN-----SAMPK0SKIK 1739
QY 759 IHKVP0K0IASYQ0-----YKNE---0ESRVTSPKVKI0PNAI0EKKKREVNNMR- 806
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1740 PF0RV--KKIM0V0Q0ASASASSAPNK0LDGKKKPTSPVKPI0P0NTEVYTRVAKNADS 1797
QY 807 ---RVVSPMDLANS0FLPELSEDSGFKDLNANSNNNR-----PSSFTPL 852
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1798 NLN0ERVS0DKND--SKK0NLKNNSKD--FND---KLPPNEDRV0GSAF0DSPHIHYPI 1849
QY 853 ---STKNVLSNIDNDPVNEPPEPKSYAEIRNARLSANKA0PNA0APLP0Q0R0S 905
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1850 EGT0PCF0RND0SLSD0DDDDV0DLSREK--AELRKAKENKSEAKVYSHTEL-----T 1901
QY 906 STRNSNRKVR0FRVPT0EIRRTSSALAPCDMYNDI0FDEFGAGSKPTJKAESMKTPLPSMD 965
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1902 SNO0SANKT0AIK0P---INR-----GQPKPII0K0--STFP0SS 1937
QY 966 K0DVYRII0NAKKGV0DEYI0NAKLVD0K0P-----KKN0SV--RD 1002
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1938 K0IPDR-----GAATDEKL0NFAI0NTPVCFSHNSSLSSLSLSDI0ENNKNENEPKETE 1991

```

```

QY 1003 PEDRYEE-----L00TASIHNAITDSIYGRPD0ISTDMLPYL 1040
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1992 PPD0SG0E0PK0P0ASG0YAPK0SFHVEDT0PVCF0SRKSSLSISIDSE-----D0LIQ0E0ISSA 2046
QY 1041 SDELK0PPTALLSADRLEF0E0VHPLRNSVYVHPGAGATNSSMLPEP0FELINS0PARN 1100
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2047 MPK0KKKP-----SRLK0DNEKH0SPRNM0GILL--G0DLTL0DKI0RDP0SEH0GLSPD-- 2095
QY 1101 V0NNS0NVAI0SGNASTI0SN0LDM0NFD0ATIG0KI0E0BP0ASKANTV0GDD0GLASAP 1160
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2096 -SENEDWKAI0EGANSI-----VSSLH0MAAA0CLSR0-ASSDS0SILSLKSGISL0GSP 2147
QY 1161 TPRTPKKESIS0KPAKLSASPRKSPIKIGSPVRYIKKNGSIAGIEPIPKATHPK0KSF 1220
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2148 FHUTP0OE-----KRP-T0NKGR-----ILKPEK0STLE--TK0IES0KGI 2188
QY 1221 Q0NEI0SNH0KVRD0GIS0PSSGSEH0Q0HP-----SNV0SP0YTDA0T0V0PDEN 1268
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2189 K0GK0K0YK0SLIT0KVNS0EIS0G0MK0P0L0ANMP0IS0RGTMIH0LGVN0SS0STS0P0SK 2248
QY 1269 K 1269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2249 K 2249

```

```

RESULT 13
US-10-022-819-2
; Sequence 2, Application US/10022819
; Publication No. US20030027166A1
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Antonette C. P.
; OLSEN, Sheri J.
; LAWRENCE, Tammy
; ANGELLY, Tracy S.
; RABIN, Mark B.
; TITLE OF INVENTION: CODING SEQUENCE HAPLOTYPE OF THE HUMAN
; BRCAL GENE
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan Lewis & Bockius LLP
; STREET: 1111 Pennsylvania Avenue
; CITY: Washington DC
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible.
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/022,819
; FILING DATE: 22-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/074,452
; FILING DATE: 1998-05-06
; ATTORNEY/AGENT INFORMATION:
; NAME: <Unknown>
; REGISTRATION NUMBER: <Unknown>
; REFERENCE/DOCKET NUMBER: 044921-5049-01-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-739-3000
; TELEFAX: 202-739-3001
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO

```

```

: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: HOMO SAPIENS
: STRAIN: BRCAL
: HAPLOTYPE: OM14
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: 17
: MAP POSITION: 17q21
: SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-022-819-2

```

```

Query Match      3.18; Score 268.5; DB 9; Length 1863;
Best Local Similarity 18.68; Pred. No. 3e-06;
Matches 333; Conservative 243; Mismatches 604; Indels 607; Gaps 81;

```

```

18  OLOPOSSASAFNSPTKPLNFPRTNSKPSLDPNSSD-----TYSSEDOEGKREKEDT 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  170 RIQPKTSVYI-----ELGSDSEDTYVKATYCSVGDDQLQITPGT 212

   73 AFQTSFDR-----NFDLNSIDIOOTIQHO-----OQOPQOQOOLQOTDN 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  213 RDEISLDSAKKACEF---SETDVTNTEHQPSNNDLNTTEKRAERHREKYQG--SSVSN 268

   113 NLIDFSGTQMT-----STLDLTKONPTVDKV---NENHAPTYINTSPKSKTKKAT- 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  269 LHVPCGNTHTASSLQHNSSLLTKDRMNEYKAEFCNKSQOPGLARSOHNRMAQSKETC 328

   163 -PKAPKVAFTVNTPEINHHPDNVVEEDQSOOK-----EDSEVERPLIQ----- 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  329 NDRKTPS-----TEKKYDLNADPLCEKEMKOKLPCSEMPROTEDEVPTWILNLSIOKV 382

   207 HQMKDPSQFNYSDE-----DTNASVPTPLPTTKPTFAQLLNKNEVNSEPE 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  383 NEW-----FSRSDELGSDSDHGESESNAKV-----ADVLVDLNEY----- 419

   255 ALTMKALKRENFSLIDEKYNLYLSPNNNNKSNVSDMSHLQNLDAASKNTENIHN 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  420 -----DEYSGSS--EKIDLLAS-----DPR-----BALICK--SERVHS 449

   315 LSLFALKPKNDIENPLNLTNADISLRSSGSSQSLQSLRDNRYLESVPSPKKVNPGL 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  450 -----KVESNIEDIK-----FGKTYRKASLPLNLSHTENLI----- 482

   375 SLNDGKIGFSOE---VVESSLPRDLSRDKLETTKEHDAPEHNNEFIDAKSTNTKGGQL 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  483 -----IGAFVTEPOIIOERPLTNKLRKRRTSGLHP-----EDFI-----KKADLA 524

   432 VSSDHLDSFDRSYNHPQS---IINLNASASOSQSLALEKOR-----QTOEOQOTQAA 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  525 VQKTPKM--INQGTNQTQNGOVMMNTSGHENTKKGDSIONEKPNPIESLEKESAKFT 582

   485 EPEETSFSDNIKV-----QEPKSN-----LEFVVTIKKSPVASTEIK 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  583 KAEPISSISMELNLTNHNKAPKKNLRKRSSTRHIALELV--VSNLSPPNPTQLOI 641

   526 PKRFSSSILRIKED-----ETAPADIIHK--KENANSHVEDTDLALKA 571
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  642 DSCSSSEIIRKKYKNOMPVHRSRMLQMEGKEPATGAKKSNKPNQOTSKRSDTFPELK 701

   572 LNDDEESTQNTSKMSIRF-----HIDSDMKLED--SNDGREDNDQI-----SREK 618
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  702 LTNAPGSTTKCSNTSELKEFVNPSPLEPEKEKELETYKVSNNNA--EDPKDLMLGSEVLQ 759

   619 SDILNDVQSTDIIGDKYG--NSSSEITTKTLAPPNSDND-----KENSSTL----- 664
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  760 TERVESSISLVPGTQGTQESTISLEVTLGAKATEPNKCVSOGCAFENPKGLIHCS 819

   665 EDPANN--ESLQOOL--EVPATKEDDSILANSSNIAPEELTLPVVEANDYSSFN----- 715
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  820 KDNNDTQEGFYPLGHEVYHNSRETSIEMESELDAQYLQNTFFKYSKRSQSLFNSPGNAE 879

   716 DVTTFDAVSSFEESLSREHETDSKPINFISIMHKOEKQKHO--IHKVPTKQIIIASIQO 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db  880 ECATFSASHSG---SLKQO---SPKYTF--EEOKEENQGNESNIKPEVQTNITAGE-P 930

   774 YKNEQESRVTSDPKVILPAIOFKKKEEVNVMRSRVSPDMODLVNSQF--LELSDSQF- 831
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  931 VVGQDKRPVDNAKCSITKGSRCFLSSQFRGNETGILTNKIGELLONPRIRPLPFIKSFV 990

   832 -----KDL---NFANYS-----NNTNRRPSFTPLSTKNVLSNIDNPVNEPEPKSY 876
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  991 KTKCKKNLLENFEEHSMSPEREMENENIPSTVSTISNNINREN----- 1035

   877 AEIRNARLSANKAAPNAPPLPQROPSTRSNSNKKVSRFRVTFEIRRTSALACD 936
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1036 -----FKGASSNINEVSGSTNVGSSINEIGSSD 1065

   937 MYNDIFDEFGAGSKPTIKAEGBKTLPSMDKDDVKRLIAKKGVQDDEVYINAKLYDQKPK 996
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1066 --ENIQAEIGNRREPKLNA-----MLRGLVLOPEYVKOSLPSGSKKH 1105

   997 NSIVTDPEDRYEELQOTASIHNATIDSSIYGRPDSISTDMPLYL--SDELRKPTALLSAD 1055
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1106 PEL---KKOEYEEVYQT-----VNTDFSPYILSDMLEQ----- 1136

   1056 RLFEHQEVNPLRSNVLVNPGAGAATNSMLPEPDFELINSPARVNSNDSNVAISGNAS 1115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1137 -----MGSSHASQVCSERPDDL-----DDEGEIKEDT- 1163

   1116 TISFNOLDMPDDOATIGOKIOEOPASKS-----ANTVRGDDGLASAP 1159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1164 --SFAENIK--ESSAVFESKSVQRGELSPSPFTHTLAOGYRRKAKLESSEBLSS-- 1218

   1160 ETPRTPTKESISSKPAKLSSAPRSPKIKGISPVYIKKNGSIAGIEPIKATHKPKKS 1219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1219 EDEELPCQHLLEFGKVNINIPQSTRHSYV-----ATCLSKNTEBNLS 1262

   1220 FQG--NEISNHNKVRGSGISPSGSEHQOHNPSMVP--SQYTDATSTYVDENKOVQHNPR 1276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1263 LKNSLNDCSNOVILAKASOEHNHLSSETKCSASLFSQCSLELDTANTNTDPPFLIGSSK 1322

   1277 EKOQKHNHNNHNNHNNQO---KTDIRGVVDEIIPVGLQERGKLEFRVLGIKINILPDIN 1333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1323 QMR-----HQSESGVGLSDKELVSDDEBERTGILENNQ----- 1356

   1334 THKGRFTLLDNGVCHVTTPREYNNMDH--NVAIGKEFLTYAASLEFLLTLKASTYKPRG 1391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1357 -----EEQSMDSNLGEAASGCESETSVSD----- 1381

   1392 TLVEYTEKKVVKSRNRSRLGSKDILITTTKFPVTEVDTMANKFAPDGSFARCYIDLQO 1451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1382 -----CGSLSSQSDILT-----TQORDTQOHNUL-----IKLQO 1409

   1452 FEDQITGRASQFDLNCFFNEM-----ETMSNGNO-----PMKRGKPYKIAQLE 1493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1410 EMALAEAVLEQHGOPSMSYSITISDSSALDLNRPBQSTSEKAVLTQSKSEVPISO- 1467

   1494 VKMLVPRSDPREILPISIRSAYESINELNNDQNNYFEGYILHQEGGDCPTKKKPEFKMG 1553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1468 -----NPEGSLADFEVESADSTSKNKE---PGVRSRSPSKCPSLDDRRVYHSC 1513

   1554 TSLAHSEISIKTRAKINLSKVVDLIYVDKENIDKNSNRNSDYL 1600
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1514 GGLONRNP-----SQEELIKVVD--VEEOULESGPHDLTETSYL 1553

```

```

RESULT 14
US-09-734-672-4
: Sequence 4, Application US/09734672
: Publication No. US20020183268A1
: GENERAL INFORMATION:
: APPLICANT: Murphy, Patricia D.
:           Alvaras, Christopher P.
:           Allen, Antonette C.
:           Critz, Brenda S.
:           Olson, Sheri J.

```

Schelter, Denise B.
Zeng, Bin
TITLE OF INVENTION: Coding Sequences of the Human
BRCA1 Gene
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan Lewis & Bockius LLP
STREET: 1111 Pennsylvania Ave., N.W.
CITY: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/734,672
FILING DATE: 03-Dec-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/966,436
FILING DATE: 07-Nov-97
APPLICATION NUMBER: US 08/598,591
FILING DATE: 12-Feb-96
ATTORNEY/AGENT INFORMATION:
NAME: Michael S. Tuscan
REGISTRATION NUMBER: 43,210
REFERENCE/DOCKET NUMBER: 44921-5055-02-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-739-3000
TELEFAX: 202-739-3000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20020183268A1 Relevant
TOPOLOGY: No. US20020183268A1 Relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-734-672-4
Query Match 3.1%; Score 267.5; DB 9; Length 1863;
Best Local Similarity 18.8%; Pred. No. 3.4e-06;
Matches 336; Conservative 243; Mismatches 601; Indels 607; Gaps 82;

QY 18 QLOPOSSASIFNSPTKPLNFPNRSKPSLDPNSSD-----YTSDDDEKGEKKKDT 72
DB 170 RIPOKSTSYI-----ELGSDSSEDTVNAKATYCSVGDELLQITPOGT 212
QY 73 APTSTDR-----NPDLSIDIQOTIOH-----OOQPOQOQLSQTDN 112
DB 213 RDEISLDSAKKACF---SETDVTNTEHHQPSNNDLNTTEKRAAEHPEKYQG-SSVSN 268
QY 113 NLIDESFOTPMT-----STLDLTQONPTVDKY---NENHAPTYINTSPKSIIMKAT- 162
DB 269 LHYEPGCTNTHASSLQIENSILLLTKDRMVEKAKEPCNKSKOPGLARSQHNRMAKSKETC 328
QY 163 -PRASPKKAVFYTNPEIHHYPDNRVEEDOSQK-----EDSVPEPLIQ----- 206
DB 329 NDRRTES-----TEKVDLNADPLCEKKEWKKKLCPCSENPRTDEVPWITLSSIOKV 382
QY 207 HOKKDPQEQFYSD-----DTNASVPTPTPLHTTKPTFAOLKNNNNVENSEPE 254
DB 383 NEW-----FSRSDLLGSDSDSHGSESNKAV-----ADVLDLNEV----- 419
QY 255 ALIDMKLRNFSNLSIDEKVNLYLSPYNNNNKNSVSDMSHLQNLQDASKNKTNNENIH 314

DB 420 -----DEYSGSS--EKIDLLAS-----DPH-----EALICK--SERVHS 449
QY 315 LSFALKAPKNDIENPLNSLTNADISLSSGSSSLOSRLRDNRYLVESPOSPKKNVPGCL 374
DB 450 -----KSVESNIEDKI-----FGTYKKKASLPMISHVTENLI----- 482
QY 375 SLNDGIGFQDE---VYESLPRDLSRDKLETTREHAPAEHNNNPFIDANKSTNNKQGL 431
DB 483 -----ICAFYTEPOIIOERPLTNKKRRRTSGUHP-----EDFI-----KKADLA 524
QY 432 VSSDHLDSFDRSYNHTEOS---ILNLSASOSQISINALEKOR-----QTOEOBOTQA 484
DB 525 VQKTPKM--INQGTNOGEQNGVANNITNSCHENTKGSIQENKNPNPESLEKESAFT 582
QY 485 EPEEETSFSDNIKKV-----QEPKS-----LEFYKVITKKPVSATETIKA 525
DB 583 KAEPISSISNMELELNHNSKAPKKNLRKRSSTRHIALELV-VSRNLSPPMCTELQI 641
QY 526 PKREFSRIILIKMED-----ELAEPAIDHPK--KENEANSHYEDTALLKKA 571
DB 642 DSCSSSEIKKKKKNQMPVRHSRNLOLMEGKEPATGAKSKKPNEDOTSKRHDSOTFPELK 701
QY 572 LNDEESDTTONSTKMSIRF-----HIDSPMKLED--SNDGREDNDI-----SPREK 618
DB 702 LTNAPGSFTKCSNNTSELKEFYNPSPLPREKEKELETYKVSNNA--EDPKDMLSGERYLQ 759
QY 619 SDILINDYSQTSIDITGDKYG--NSSSEITTKTLAPPRSDND-----KENSKU----- 664
DB 760 TERSESSSISLVPGTGYDQESISLLEVSTLGAKAKEPNKCVCQAFAFENPKGLIHGS 819
QY 665 EDPANN--ESLQOOL--EVPHTKEDDSITLANSNLAPEELLPVEANDYSFN----- 715
DB 820 KDNNDNDEGEFYPLGHEVYHNSRETSIEMESELDAQYIQTFFKSKRQSPALFSFGNAE 879
QY 716 DVTKTFDAYSSFEESLSREHETDSKPINFISIMHKOEKOKHQ--IHKVPTKQI1IASYOQ 773
DB 880 EECATFSAHSG---SLKKQ---SPKVTF--ECEQKEENQKNESNIKIPQYVINITAGF-P 930
QY 774 YKNOESRYTSDKVKIRPAIOFKFKFVYVNSKRVYSDMDLNVSCF--LELSDSGSF- 831
DB 931 VVGQKDKPVDNAKCSINGSRFLCSQFRGNETGLTPNKGGLLQNPRIPLPPIKSFV 990
QY 832 -----KDL--NFANYS-----NNTNRPSFTPLSKVNLSDNDPNVVEPEPKSY 876
DB 991 KTKCKKNLLENFEBHSSPREKNGENIPSTVSTISNNIRENVFKE----- 1038
QY 877 AEIRNARLSANKAPNQAPLPPOROPSSSTRSNSNKRVSFRVPTFEIRRTSALACD 936
DB 1039 -----ASSNINEVGS-----STNEVGSSINIEIGSSD 1065
QY 937 MYNDIFDDFGAGSKPTIKABGKTLPSMDKDDVKRILNAKGVYODEYINAKLYDQAKK 996
DB 1066 --ENIQAEELGRRNRGPKLNA-----MLRGLVLOPEVYKOSLPESNCKH 1105
QY 997 NSIYTDPEDRYEELQOTFASINNATIDSIGRPDISISTDMPLPYL--SDELKRPPTALLSAD 1055
DB 1106 PET---KKOYEIEVYOT-----VNTDFSPYILSDMLEOP----- 1136
QY 1056 RLFMEOEYHPLRNSVYLHPGAGATNSSMLPEPDFELINSPARVNSNNSNVAISGNAS 1115
DB 1137 -----MGSSHASQVCEPTDDLL-----DDGEIKEDT- 1163
QY 1116 TISFNOLDAMPNDDOATIGOKIOEOPAKS-----ANTVRGDDGLASAP 1159
DB 1164 --SEAEINDIK--ESSAVFSKSYQKGLSRSPFTHTHLAGOYRGAKKLESSEENLSS-- 1218
QY 1160 ETPTPTPKESISSPAKLSSAPKPSPIKIGSPYRVIKKNGSISAGIPKATPKPKS 1219
DB 1219 EDEELPCGHLLFGKVNINIPQOSTRHSTV-----ATCLSKNTEENILS 1262
QY 1220 FQG--NEISNHRKYVRDGTSPSSGSEHQOHNPMSVVP--SQYTDATSTVPDENKDVQHKPR 1276

```

Db 1263 LKSLNDSCNOYLAKASQEHHLSEETKCSASLFFSSQCELEDLTANTNTODPFLIGSSK 1322
Oy 1272 EKOKKHNRHHHHHHKQ--KTDPGVYDDEIDPVGLOERGLFPRVIGIKINILPDIN 1333
Db 1323 QMR-----HQSQGVGLSDKELVSDDERGTGLENNQ----- 1356
Oy 1334 THKGRFTLLDNGVHCVTTPPEYNMDH--NVAIGKEFELTVADSLFEILTLKASYEPKRG 1391
Db 1357 -----EEOQMSDNLGEAASGCESTSVSD----- 1381
Oy 1392 TLVEYTEKVVASRNRLSLFGSKDIITTTKPEVTKTWMANKFAPDGSFARCYIDLQO 1451
Db 1382 -----CSGLSSQSDILT-----TQQRDTMOHNL----- 1409
Oy 1452 FEDQITGKASQDFLNCFNFM-----ETMNCNQ-----PMKRGPKYIAOLE 1493
Db 1410 EKAELVAVLEQHGOSQSNRYPTSISSSALDLRNEQSTSEKAVLTSQKSEYPTSQ-- 1467
Oy 1494 VKMLVYPRSDPREILPTFSIRSAVESINELNNEQNNYFEGYLHQEGGDCPTFKKRPFKLNG 1553
Db 1468 -----NPEGLSADKFEVSADSTSKNKE-----PQVERSSPSKCPSLDRWYMHSC 1513
Oy 1554 TSLAHSEISHKTRAKINLSKVVDLYVDKENIDRSHNHFSDVLL 1600
Db 1514 SCSLQNRNP-----SOEELIKVVD--VEEQLESGPHDLTETSYL 1553

```

```

RESULT 15
US-09-982-828-6
; Sequence 6, Application US/09982828
; Publication No. US20030022184A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; Allen, Antoinette C.
; Alvarez, Christopher P.
; Critz, Brenda S.
; Olson, Sheri J.
; Thudner, Denise
; Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; BRCAL Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Morgan Lewis & Bockius LLP
; STREET: 1111 Pennsylvania Avenue N. W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/982,828
; FILING DATE: 22-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/074,453
; FILING DATE: 1998-05-06
; APPLICATION NUMBER: US 08/738,691
; FILING DATE: 1997-02-12
; APPLICATION NUMBER: US 08/598,591
; FILING DATE: 1996-02-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Michael S. Tuscan
; REGISTRATION NUMBER: 43,210
; REFERENCE/DOCKET NUMBER: 44921-5053-01-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-739-3000
; TELEFAX: 202-739-3001
; INFORMATION FOR SEQ ID NO: 6:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL (om13)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-982-828-6

Query Match 3 1%; Score 267.5; DB 9; Length 1863;
Best Local Similarly 18.8%; Pred. No. 3,4e-06;
Matches 336; Conservative 243; Mismatches 601; Indels 607; Gaps 82;

Oy 18 QLDQSSASIFNSPYKPLNFPRTNSKPSLDPNSSSD-----TYTSEODQEKKEKEDPT 72
Db 170 RIQPKTSYVI-----ELGSDSSSDTYNKKATYCSVGQQLLQITPGQT 212
Oy 73 APQSTFDR-----NFDLNSIDIQOTIOH-----QQPQQQQQLSQTGN 112
Db 213 RDEISLDSAKKACEP-----SETDVTNTEHHQPSNNDLNTTEKRAERHBEKYQG--SSVAN 268
Oy 113 NLIDFSPQTPMT-----STDLTKQNPYDKV--NENHAPTYINTSPKNSIMKAT- 162
Db 269 LHYEPCGTTTHASSLOHENSLLITKDRMNAVEKAEFCNKSQDGLARSOHNRAGSKETC 328
Oy 163 -PKASPKVAFTVTPNPEIHHPDNRVEEDQSOQ-----EDSVEPPLIQ----- 206
Db 329 NDRRTPS-----TEKKVLDLNDPLCEKKEWMKQKLPCEENPRDITDVMWITLNSIQKY 382
Oy 207 HQMKDPSQFNYSDE-----DTNASVPTPTPLTTKPTFQQLLNKNEVSEDE 254
Db 383 NEW-----FSRSDDELGSDDSHDGESESNKV-----ADVLDVINEV----- 419
Oy 255 ALIDMKLKEFNFSNLSDLEKVNLYLSPYNNNSKNVSDMSHLQNLQDSKKNKTENIN 314
Db 420 -----DEYSGSS--EKIDLLAS-----DPI-----EALICK-SRVHS 449
Oy 315 LSFALKAPKNDIENPLNLSLTNADISLRSSGSSQSSLSQSLRNDNRVLESVPGSKVNPGL 374
Db 450 -----KSVESNIEDKI-----PGKTYRKKAASLPNLSHYENLI----- 482
Oy 375 SLNDGIKGFSDE---VESLPLPDLSRDKLETTKEKDAPEHNNEINTDAKSTYTKGQLL 431
Db 483 -----IGAFTVEPOLIIOERPLTNKLRKRRTPSGLHP-----EDFI-----KKADLA 524
Oy 432 VSSDDHLDFDRSYNHTEOS--ILNLLNSASQSOISLNLKOR-----QTOBOEQTOAA 484
Db 525 VQKTPEM--INQNTOTEDNGOVAMNTTNSGHEKTKGDSIQNEKNPNPLESLEKEAFPT 582
Oy 485 EPEETSFSDNIKVK-----QEPKSN-----LEFVKYTIKKEPVASATEIKA 525
Db 583 KAPISSISNMLELINHNSKAPKRNRLRKRSSTRIHILELY--VSRLSPNCELEJO 641
Oy 526 PKREFSSRLIRIKNED-----ELAEPADLHPK--KENENSHVEDDLDALLKKA 571
Db 642 DSCSSSEETIKKKRYNOMPVYHSHRNLOLMGKKEKATGAKKSNKPNEDTSCRHSDTPELK 701
Oy 572 LNDDESDTTONSTKMSIRF-----HIDSMDKLED--SNDGDRDNDNDI-----SRFK 618
Db 702 LITAPGSFTKCSNTSLKEFVNPSPREKKELELYVYKSNNA--EDPRDLMSGRVQLQ 759
Oy 619 SDILNDVSOQSLIIGDKYG--NSSSEITTKTLAPPSSDND-----KENSKL----- 664
Db 760 TEKSVSSSISLVPGTGYQESISLLEVSTIGKATETPNKVCQCAARENPGILHGCS 819
Oy 665 EDPANN-ESLQOOL--EVPHTKEDDSILANSSAIAPPEELITLPVYVANDYSSFN----- 715

```


Db 820 KDNRRNDTEGFKYPLGHEVNHRSRETSIEMESELDQAQYLTQNTFKVSKROSFALFSPGNAE 879
QY 716 DVTKTFDAYSFEESLSREHETDSKPIINFISIMHKOEKKHQ--IHKVPTKOIIASYQ 773
Db 880 EECATFSAHG---SLKKQ---SPKYTF--ECEQKEENQKNESENIKPVQTVNITAGF-P 930
QY 774 YKNEOESRYTSDVKIIPNAIQFKFKFEVNVMSRRVSPMDDDLNVSOQ-LPELSEDSGF- 831
Db 931 VVGOKKQPVDMNAKCSIKGSRFCSSQFRGNENGLITPNKHGLQNPYRIPPLFPKSFV 990
QY 832 ----KDL---NFANTS-----NNTNRRSFTPLSTKNVLSNIDNDPVNVEPPEPKSY 876
Db 991 KTCKCKNLEENFEHESMSPEREGNENIPSTVSTISRNIRENVFKE----- 1038
QY 877 AEIRNARLSANKAPQAPPLPQROPSSSTRNSNKRVSFRVPTEIRTSALAPCD 936
Db 1039 -----ASSNINEVS---STNEVGSSINEIGSSD 1065
QY 937 MYNDIFDDFGAGSKPTIKAGMKTLPSMDKDYKRIINAKKGYTODEYINAKLVQCKPKK 996
Db 1066 --ENIQAEIGRRNGPKLNA-----MLRLGVLPQPEVYKOSLPGSCKH 1105
QY 997 NSTVTPEDRYEELQOTASTIHNAITDSIYGRPDSISTDM.PYL-SDELKKPTALLSAD 1055
Db 1106 PEI---KKOYEYEVQF-----VNTDESPYLLISDNLEQ----- 1136
QY 1056 RLFMEQEVHPLRNSVLYVHFGAGATNSMLPEPPFELINSPARVNSNDSNVAISGNAS 1115
Db 1137 -----MSSHASQVCEPDPDL-----DDGEIKEDT- 1163
QY 1116 TISFNQLDNMFDDQATIGOKIOEOPASKS-----ANTVRGDDDLASAP 1159
Db 1164 --SFAENDIK--ESSAVFSKSVQKGLSRSPFTHTHLAGYRRGAKKLESSEENLSS-- 1218
QY 1160 ETPRTPTKKSISSKPAKLSASAPKSPKIKIGFVRIKKNGSTAGIEPIPKATHPKPKS 1219
Db 1219 EDEELPCFOHLLFGKVNINIPSOSTRHSTV-----ATECLSKNTEENILS 1262
QY 1220 FQG--NEISNHNKVRDGISSPSSGEHQHNPMSVVP--SOYDATSTVPDENKDVQHKPR 1276
Db 1263 LKNSLNDCSNOVILAKASQEHNLSEETKCSASLFSQCSLELITANTNTQDPPLIGSSK 1322
QY 1277 EKOKOKHHNHHNHHNQ---KTDIPGVVDEIIPVGLQERGLFFRVLGIKNINILPDIN 1333
Db 1323 QMR-----HQSQOGVGLSDKELVSDDEERGTLGEENNQ----- 1356
QY 1334 THKGRFTLIDNGVHCYTPPEYMMDH--NVAIGKEPELTVADSLFILTAKASYEKP RG 1391
Db 1357 -----EOGMDSNLGEAASGCESETSVSED----- 1381
QY 1392 TIVEYTEKKVYKSRNRLSRLFGSKDITTTKVFTEVKDWTWANKFAPDGSFARCYIDLQ 1451
Db 1382 -----CSGLSSQSDILT-----TOQRDTMQHNL-----IKLQ 1409
QY 1452 FEDOITGKASQFDLNGFNEM-----ETMSNGNQ-----PMKRGRPKYIAQLE 1493
Db 1410 EMAELEAVLEHQSPNSVPIIISDSALEDLRNPQOSTSEKAVLTSQKSSEYPIQ-- 1467
QY 1494 VKMLYVPRSDPREILPTISINAVESINELNNEQNNFEGYLHOEGDCPIFKKREFKLMG 1553
Db 1468 -----NPEGLSADKFEVSADSTSKNKE-----PGEVRSSPSKCPSLDDRWYMHSC 1513
QY 1554 TSLLAHSEISHKTRAKINLSKVVDLIYVDKENIDRSNHRNPSDVL 1600
Db 1514 SGLSLQNRNTP---SOBELIKYVD---VEQQLSESGPHDLTETSTYL 1553

Search completed: March 17, 2003, 12:34:17
Job time : 93.588 secs

This Page Blank (uspto)

Shears, Beverly

From: Devi, Sarvamangala
Sent: Tuesday, March 25, 2003 7:02 AM
To: Shears, Beverly
Subject: RE: 09/964,858

Good morning Beverly:

Sorry Beverly. I have not received sequence search report for this case. I would appreciate if could please perform a sequence and an interference search for SEQ ID NO: 1 or print out the sequence report from the March 17th search, if it was saved. Thanks a lot.

-----Original Message-----

89182

Sh ar s, Beverly

From: Devi, Sarvamangala
Sent: Friday, March 14, 2003 4:38 PM
T : Shears, Beverly
Subject: 09/964,858

Hello Beverly:

Would you please perform a sequence and an interference search for SEQ ID NO: 1 and a fragment comprising amino acid residues 1-263 of SEQ ID NO: 1 in case 09/964,858?

Thanks.

S. DEVI, Ph.D.
AU 1645
CM1-7E15

7E12
mb

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

This Page Blank (uspto)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:19:45 ; Search time 9.28075 Seconds
(without alignments)
5839.012 Million cell updates/sec

Title: US-09-964-858-1_COPY_1_263
Perfect score: 1386
Sequence: 1 MNSPKLPLDKHSHLQLQ.....NKNNEVNSEPALTDMLKR 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

arched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	158.5	11.4	1709	5	Q95PH5
2	146	10.5	542	5	Q96600
3	145.5	10.5	2454	5	Q8R2G3
4	143.5	10.4	1721	5	Q8SS00
5	142.5	10.3	2678	5	Q9NDS4
6	142	10.2	2843	5	Q963J5
7	138.5	10.1	734	5	Q8T1A0
8	138.5	10.0	1166	2	Q86489
9	137	9.9	730	5	Q8R867
10	137	9.9	756	11	Q9QY60
11	137	9.9	947	5	Q15739
12	135.5	9.8	2472	5	Q8R2M5
13	134.5	9.7	1162	4	Q96JX9
14	134.5	9.7	1171	2	Q9KWX6
15	134	9.7	1141	16	Q99W46
16	134	9.7	1141	16	Q932F7

17	133.5	9.6	800	5	Q96902	Q96902 dictyosteli
18	133	9.6	628	5	Q960L0	Q960L0 drosophila
19	133	9.6	767	5	Q9V9Z6	Q9V9Z6 drosophila
20	131.5	9.5	1179	12	Q91L98	Q91L98 white spot
21	131.5	9.5	1180	12	Q8VAS9	Q8VAS9 white spot
22	131.5	9.5	2439	5	Q9VMS2	Q9VMS2 drosophila
23	130.5	9.4	1140	5	Q8T1T0	Q8T1T0 drosophila
24	129.5	9.3	803	5	Q9BLH2	Q9BLH2 halocynthia
25	128	9.2	548	5	Q8T1H3	Q8T1H3 dictyosteli
26	127.5	9.2	1335	2	Q93635	Q93635 staphylococ
27	127.5	9.2	1542	5	Q9VE07	Q9VE07 drosophila
28	127.5	9.2	1789	5	Q8R2E3	Q8R2E3 dictyosteli
29	127	9.2	736	3	Q74500	Q74500 schizosacch
30	127	9.2	816	5	Q8R2N6	Q8R2N6 dictyosteli
31	127	9.2	1338	5	Q77306	Q77306 plasmodium
32	127	9.2	1483	5	Q8T273	Q8T273 dictyosteli
33	126.5	9.1	601	3	Q9C1I3	Q9C1I3 candida alb
34	126.5	9.1	660	5	Q94485	Q94485 dictyosteli
35	126	9.1	1593	5	Q8T1T7	Q8T1T7 dictyosteli
36	125.5	9.1	838	5	Q9Y0C9	Q9Y0C9 dictyosteli
37	125	9.0	641	5	Q61640	Q61640 drosophila
38	125	9.0	646	16	Q9KM25	Q9KM25 vibrio chol
39	125	9.0	720	5	Q23847	Q23847 dictyosteli
40	125	9.0	722	5	Q9VEK4	Q9VEK4 drosophila
41	125	9.0	1545	5	Q9YVY1	Q9YVY1 drosophila
42	124.5	9.0	556	5	Q95RC6	Q95RC6 drosophila
43	124.5	9.0	773	16	Q9SR04	Q9SR04 mycoplasma
44	124.5	9.0	811	5	Q9VT40	Q9VT40 drosophila
45	124.5	9.0	870	13	Q9PTB3	Q9PTB3 colurnix co

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	1709 AA.
Q95PH5			
AC Q95PH5:			
DT 01-DEC-2001 (TREMBLrel. 19, Created)			
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE Histidine kinase DhkL.			
GN DHK.			
OS Dictyostelium discoideum (Slime mold).			
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.			
OX NCBI_TaxID=44689;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=AX4;			
RA Anjard C., Loomis W.F.;			
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AF362373; AAK54092.2; -			
KW kinase.			
SQ SEQUENCE 1709 AA; 192603 MW; F0A91C505D5DE178 CRC64;			
Query Match	11.4%	Score 158.5;	DB 5; Length 1709;
Best Local Similarity	26.4%;	Pred. No. 0.0078;	
Matches 71; Conservative 34; Mismatches 127; Indels 37; Gaps 8;			
QY 2 MNSPKLPLDKHSHLQLP-----QSSASIFNSPTKRLNPRTRNSKSLDPSNSSDPT 55			
Db 456 MNSPSSHTP---NSPMIFQPIVSNICNSGSGSNNSPHIIDN-----NNSNQOO 502			
QY 56 YTSQDQDEKGEKKDPAFTQTSFRNFDLNSIDI---QDTIQHQOQPPQOQOOLQOTDN 112			

DT 01-MAY-2000 (TREMBLrel. 13, created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Cg1/xap80 protein.
GN Cg1/xap80.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Plitzer W., Zhao W., Herman G.E., Rosenthal A.;
RT "Comparative sequence analysis of the mouse Mdm locus and the
RL corresponding region of human xq28.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AF125313; AAF22120.1; -;
DR InterPro: IPR002965; P. rich-extensn.
DR PRINTS: PR01217; PRICH-EXTENS.
CC SEQUENCE 756 AA; 81963 MW; 8F5B672B378A5943 CRC64;

Query Match	9.98;	Score 137;	DB 11;	Length 756;
Best Local Similarity	25.08;	Pred. No. 0.099;		
Matches 65;	Conservative 28;	Mismatches 125;	Indels 42;	Gaps 9

```

0Y 5 PSLIRIDKHNHLOIOPSS-SASIFNPRTPLNEPRTNKPSIDPMSSDYTSEDDE 63
    1 1 1 : : 1 1 1 1 : : 1 1 : : 1 1 : : 1 1
Db 338 PAKRLGRLNLNORPORSOSILANLVSSVKSROGHLISALRTGT-GRSPRTREMLSS 366
    1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1
0Y 64 KXKEKKDTATQTSFDNRF-DLDSIDIOQTIONOOROROOOOLSTDNMLDESSQT 122
    1 : : 1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1
Db 397 PGLPOOSFT-P-QUSILRSLPTSMLSOOOOROOOOROOOOROOOOROOOANSIFK 455
    1 : : 1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1
0Y 123 PMS-----TLDTLKNFTYDVKNVENHAP-TYUHT-----SPNK----- 155
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 466 PMSSTOQKTLISMIOOOLSSSEPARERTFTSMTKRLSHNVSEPRQKAMASMTSRÖS 515
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0Y 156 ---SLMKATGRKASPKYAFVTVDEI-----NNYRDKRVBEEDOSOQKEDSVRL 204
    1 1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1
Db 516 SLILNLRPQATGRANRASAATATATLÖJNNHOONNNHOONNNHOONNNHOONNNHOQ 575
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0Y 205 IONOKRDRSOFNYSDDETINA 224
    1 1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1 : : 1 1
Db 576 OOHOOOONOOOOROOOORPOSS 595
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

```

SEQUENCE 11
15739
ND 015739 PRELIMINARY; PRT; 947 AA.
AC 015739;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DI 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE Hele (Fragment).
GN HELE.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=4689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Loomis W.F., Iranfar N.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases
DR EMBL; AF019981, AAB70840.1; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF001176; SNF2_N. 1.
DR SMART; SM00487; DEXDc 1.
FT NON_TER 947 947
SQ SEQUENCE 947 AA: 109535 MW; EC2A44484368B5DD CRC64;

```

Query Match	9.98;	Score 137;	DB 5;	Length 947;
Best Local Similarity	21.18;	Pred. No. 0.13;		
Matches	52;	Conservative	45;	Mismatches 108;
			Indels	42;
			Gaps	8

```

0Y 17 L0L0P0SSASIFN-----SPTKPLNPRNSKPLDPNSSPTJYSEODQEKGE 68
Db 28 I0T0PDLFISSNIFESVDM5APRPSQKKOKITISNNNNLNNNTNNTNNNNNNKNNKN 87
0Y 69 KKD-TAQT5FDRNFDLNSIDIOOTLOH0Q00PQ0Q0LS-0TDNNLIDFESFQTPMTS 126
Db 88 NKNKNKNKNNNNNKNNFDESSFEFEEDDHARLEFEKYISIDDEEIIDIDIPF----- 141
0Y 127 TLDLTKQPTDKXNENHAPYINTSPNSKSIIMKATPKASPKKVAATYVNPRLHHYDNR 186
Db 142 -VDSENCKITIDESKDEREDDYEDNENIRIKQITISITTSITTSIP----- 191
0Y 187 VEED0S0Q0ED5VEPPLIOH0WK6PSQFYS-EDTNA5VPRPLPHTTKRFAQLN 244
Db 192 -----QK0Q0QNTPRHSP-----SPQSRKTRLOOTPLTRQTPR--QTSPSSQNNNN 237
0Y 245 KNNFNYS 251
Db 238 NNNNNNN 244

```

RESULT 12	
Q8T2M5	
ID Q8T2M5	PRELIMINARY;
	PRT; 2472 AA

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 278.5 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:JAX.
RA Gloeckner G., Eichinger L., Szafarski K., Pachebat Y., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guiso R., Kumpf K.,
RA Tünggeli B., Cox E., Quail M.A., Platzer M., Rosenthal A., Nogel A.A.:
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115393; AAL92295.1; -.
SQ Hypothetical protein.
SQ SEQUENCE 2472 AA: 278497 MW: 30CCE7157DA008A7 CRC64;

[illegible]

RESULT 13
Q96JK9
ID Q96JK9 PRELIMINARY; PRT; 1162 AA


```

OY 58 SEDQEKKEKKDPAFQTSFDRNFDLNSIDIOQTIOHQOQOPQOQOQLSQTNNLIDE 117
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 ILEKKAPNTNNDVTKPSTSEIQTKPTTQESTNIENSQOPPTP-----SKVDNQVTDA 206
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 118 FSTQTPMTSLDLTKONPTVDKNE--NNAPTYINTSPKKSIMKKATPKASPKKVAFTV 174
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 TNPKEPVNSKEELKNP--EKLKELVENDNNTDRSTKPVATAPTSVAPKRLNAKMRFAV 264
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 175 TNPETHYDNRVE---EDOSQOKEDSVPEPLIOHQWKDPSQFNVSDEDTNNA SVPTP 230
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 265 AQPA--VASNNVNDLITVKQITKVGDKDNVAAAHGKD-----LEYDTEFTI----- 312
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 231 PLHTTKPTFAQLLNKNNEVNSEPEALTD 258
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 ---DNKYKKGDTMTINYDKNVI PSDLTD 337
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: March 17, 2003, 12:27:30
 Db time : 25.2807 secs

This Page Blank (uspto)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd..

OM protein - protein search, using sw model

Run on: March 17, 2003, 12:13:06 ; Search time 65.6274 Seconds
(without alignments)
3378.605 Million cell updates/sec

Title: US-09-964-858-1
Perfect score: 8631
Sequence: 1 MNSTPSKRLPIDKSHLQLQ.....WVNLMLQDQDQDQDQSSSQ 1664

Scoring table: BLIUSUM62
Gap 10.0 , Gapext 0.5

Archived: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8631	100.0	1664	20	AAW99462	C.albicans alpha-1
2	8631	100.0	1664	23	AAU79331	Candida albicans 1
3	8631	100.0	1664	23	AAE19799	Candida albicans 1
4	1211	14.0	236	20	AAW99456	Amino acids 218-45
5	1211	14.0	236	20	AAE19800	Candida albicans 1
6	382.5	4.4	3257	22	ABR67502	Drosophila melanog
7	375	4.3	10182	23	ABP38314	Staphylococcus epi
8	348.5	4.0	3696	23	ABP40235	Staphylococcus epi
9	338	3.9	5024	22	AAE29335	S.epidermidis ope
10	336	3.9	6815	22	ABR68811	Drosophila melanog

11	333.5	3.9	3111	22	ABR60327	Drosophila melanog
12	327	3.8	6281	22	AAU37403	Staphylococcus aur
13	325.5	3.8	2768	22	ABR68397	Drosophila melanog
14	314	3.6	4498	22	ABR58595	Drosophila melanog
15	311.5	3.6	1803	22	ABR65391	Drosophila melanog
16	310.5	3.6	1468	22	ABR62991	Drosophila melanog
17	309.5	3.6	2437	22	AAU34338	Staphylococcus aur
18	307.5	3.6	2519	22	ABR16636	Novel human diagno
19	303.5	3.5	1970	22	ABR64827	Drosophila melanog
20	303	3.5	2346	22	ABR65519	Drosophila melanog
21	300	3.5	4134	20	AAV31946	Plasmodium falcipa
22	298	3.5	2025	22	AAU34207	Staphylococcus aur
23	298	3.5	3158	22	AAU37018	Staphylococcus aur
24	297	3.4	2951	22	ABR60291	Drosophila melanog
25	295.5	3.4	1279	22	AAE85047	S.epidermidis ope
26	292.5	3.4	3080	22	ABR64877	Drosophila melanog
27	289.5	3.4	1852	22	ABR67170	Drosophila melanog
28	289.5	3.4	2344	22	AAU37120	Staphylococcus aur
29	287	3.3	5533	22	ABR65772	Drosophila melanog
30	287	3.3	5560	22	ABR71160	Drosophila melanog
31	286.5	3.3	1420	22	ABR63410	Drosophila melanog
32	285.5	3.3	1979	21	AAE18171	Plasmodium falcipa
33	284.5	3.3	1095	22	AAE83030	S.epidermidis ope
34	284.5	3.3	2663	22	AAE39097	Human polypeptide
35	282.5	3.3	2633	22	ABG05505	Novel human diagno
36	282.5	3.3	2688	22	AAE40883	Human polypeptide
37	281.5	3.3	3238	22	ABR71715	Drosophila melanog
38	280	3.2	2368	22	AAU34139	Staphylococcus aur
39	280	3.2	2368	22	AAU36796	Staphylococcus aur
40	279	3.2	2779	22	ABR62371	Drosophila melanog
41	276.5	3.2	1400	22	AAE83348	AAE-2 protein sequ
42	276.5	3.2	1863	18	AAE23285	Human breast and o
43	276	3.2	2194	22	AAE40114	Human polypeptide
44	274.5	3.2	1712	22	ABR60536	Drosophila melanog
45	274	3.2	1177	23	ABP40261	Staphylococcus epi

ALIGNMENTS

RESULT 1	
AAW99462	AAW99462 standard; Protein: 1664 AA.
AAW99462	
08-JUN-1999	(first entry)
C.albicans alpha-INTP protein.	
Integrin-like motif; vaccine; immune response; antibody; inhibition;	
adhesion; endothelial cell; pathogenesis; infection; probe.	
Candida albicans.	
US5886151-A.	
23-MAR-1999.	
03-MAY-1996;	96US-0642846.
03-MAY-1996;	96US-0642846.
03-MAY-1996;	96US-0642846.
(MINU) UNIV MINNESOTA.	
Bendel CM, Gale CA, Hostetter MK, Kendrick K, Tao NJ;	
WPI: 1999-242618/20.	
N-PSDB: AAX25885.	
New isolated Candida albicans protein with integrin-like motifs	
Examples: Column 13-14; 21pp; English.	

XX This sequence represents the *Candida albicans* alpha-INT1 protein which contains integrin-like motifs. The protein was used to derive peptides CC AA099456-W99461 used for producing vaccines for stimulating an immune response. The antibodies can inhibit the adhesion of *C. albicans* to cells, particularly endothelial cells. This blocking activity of the CC adhesion to cells can reduce or prevent subsequent events in the CC pathogenesis of invasive candidal infection.

XX Sequence 1664 AA:

Query Match 100.0%; Score 8631; DB 20; Length 1664;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSTPSKLLPDKSHSHLOLOPOSSASIFNSPTKPLNPRNPKSLDPNNSDPTTSEQ 60
 D 1 MNSPTSKLLPDKSHSHLOLOPOSSASIFNSPTKPLNPRNPKSLDPNNSDPTTSEQ 60
 QY 61 DOEKGEEKKDTAFOTSFDRNFDLNSIDTQOTIOHQOQPOQOQOQLOSTQDNMLDEFSF 120
 D 61 DOEKGEEKKDTAFOTSFDRNFDLNSIDTQOTIOHQOQPOQOQOQLOSTQDNMLDEFSF 120
 QY 121 QTPMTSTLDLKONPTVKNVNNHAPYINTSPNKSIMKKAATPKASPKKVAFTVTNPEIH 180
 D 121 QTPMTSTLDLKONPTVKNVNNHAPYINTSPNKSIMKKAATPKASPKKVAFTVTNPEIH 180
 QY 181 HYPNRYVEEDSOOKEDSVPRPLQHQWKDPSQFNYSDEDTNNAVPRPTPLHTTKPTFA 240
 D 181 HYPNRYVEEDSOOKEDSVPRPLQHQWKDPSQFNYSDEDTNNAVPRPTPLHTTKPTFA 240
 QY 241 QLLKKNVNEPEALDMLKRENFNLSLDEKYNLYLSTPTNNNSKNVSDMOSHLONL 300
 D 241 QLLKKNVNEPEALDMLKRENFNLSLDEKYNLYLSTPTNNNSKNVSDMOSHLONL 300
 QY 301 QDASKNKTENINHLNLFALAPKNDIENPLSLTNADISLRSSGSSGSSLOSLRDNKRYL 360
 D 301 QDASKNKTENINHLNLFALAPKNDIENPLSLTNADISLRSSGSSGSSLOSLRDNKRYL 360
 QY 361 ESVGSPKVPNGSLANDGKGFSDVEVESLPRDLSDKLETTKEHAPENHNENFIDA 420
 D 361 ESVGSPKVPNGSLANDGKGFSDVEVESLPRDLSDKLETTKEHAPENHNENFIDA 420
 QY 421 KSTNTNGQLLVSSDDHLDSDFRSYNHTFOSILNLSASOSISLNLLEKORQOEBQ 480
 D 421 KSTNTNGQLLVSSDDHLDSDFRSYNHTFOSILNLSASOSISLNLLEKORQOEBQ 480
 QY 481 TQAAPEEETSFSDNIKVKQEPKSNLEFVKYTIKKEPVSAFEIKAPKREFSRIIRKNE 540
 D 481 TQAAPEEETSFSDNIKVKQEPKSNLEFVKYTIKKEPVSAFEIKAPKREFSRIIRKNE 540
 QY 541 DEIAEPADIHKKENEANSHVEDTALLKALNDEESDTTQNSTKMSIRFHIDSWKLE 600
 D 541 DEIAEPADIHKKENEANSHVEDTALLKALNDEESDTTQNSTKMSIRFHIDSWKLE 600
 QY 601 DSNNGDEEDNDISRFEKSDILNDVSOITIGDKYSSSEITTKTLAPRSDNKNEN 660
 D 601 DSNNGDEEDNDISRFEKSDILNDVSOITIGDKYSSSEITTKTLAPRSDNKNEN 660
 QY 661 SKSLEDPANNESSLOOQLEVPHTKEDDSILANSSNIAPEELTPVVEANDVSSFDVTKT 720
 D 661 SKSLEDPANNESSLOOQLEVPHTKEDDSILANSSNIAPEELTPVVEANDVSSFDVTKT 720
 QY 721 FDAYSFEESLSRHEHDSKPIINFISIMHKEQOKKHQIHKVPTKQIILASYOQKNEDES 780
 D 721 FDAYSFEESLSRHEHDSKPIINFISIMHKEQOKKHQIHKVPTKQIILASYOQKNEDES 780
 QY 781 RVTSDKAKIPAIATQPKFKEVNVMSRRVYSPDMDDLNVSQLPELSEDSGCKDLNFANYS 840
 D 781 RVTSDKAKIPAIATQPKFKEVNVMSRRVYSPDMDDLNVSQLPELSEDSGCKDLNFANYS 840
 QY 841 NNTNRPSFTPLSTKNVLSNTDNDPNVEPEPKSYAIRARRLSANKAPNQAPPLPP 900
 D 841 NNTNRPSFTPLSTKNVLSNTDNDPNVEPEPKSYAIRARRLSANKAPNQAPPLPP 900

DB 841 NNTNRPSFTPLSTKNVLSNTDNDPNVEPEPKSYAIRARRLSANKAPNQAPPLPP 900
 QY 901 QROPESTRSNKRYRFRFPFEIRRTSALAPCDMYNDIDPDGAGSKPTIKAEKMT 960
 DB 901 QROPESTRSNKRYRFRFPFEIRRTSALAPCDMYNDIDPDGAGSKPTIKAEKMT 960
 QY 961 LPSMDKDVKRIILNAKKGVYODEYINAKLVQKPKKNSIVTDPEDRYEELQOTASIHNT 1020
 DB 961 LPSMDKDVKRIILNAKKGVYODEYINAKLVQKPKKNSIVTDPEDRYEELQOTASIHNT 1020
 QY 1021 IDSSYGRPDSITSMILYLSDELKKRPPALLSARLMEQVHNLRSNVLVHAGAA 1080
 DB 1021 IDSSYGRPDSITSMILYLSDELKKRPPALLSARLMEQVHNLRSNVLVHAGAA 1080
 QY 1081 TNSMILPEPDEFLINSPARNYSNNSDNYAISGNASTISFNOLDMNFDOATIGQKIQOP 1140
 DB 1081 TNSMILPEPDEFLINSPARNYSNNSDNYAISGNASTISFNOLDMNFDOATIGQKIQOP 1140
 QY 1141 ASKSANTYRGDDGLASAPETPRPTKKESISSKPAKLSSASPRKSPIKIGSPVRVIRKN 1200
 DB 1141 ASKSANTYRGDDGLASAPETPRPTKKESISSKPAKLSSASPRKSPIKIGSPVRVIRKN 1200
 QY 1201 GSIAIEPPIKATHKPKSFGSNETISNKKVPRDGGISPSGSGPHOONHNSWSPSQYTA 1260
 DB 1201 GSIAIEPPIKATHKPKSFGSNETISNKKVPRDGGISPSGSGPHOONHNSWSPSQYTA 1260
 QY 1261 TSTVDEKNKDYQHKPREKOKNHNHNNHNNHKKQTDIPGVVDEITPVGLQERKLFPR 1320
 DB 1261 TSTVDEKNKDYQHKPREKOKNHNHNNHNNHKKQTDIPGVVDEITPVGLQERKLFPR 1320
 QY 1321 VLGININLPRDINTHKGRTLLDNGVNCVTTREYNMDHNVATKEREELVYADSLETL 1380
 DB 1321 VLGININLPRDINTHKGRTLLDNGVNCVTTREYNMDHNVATKEREELVYADSLETL 1380
 QY 1381 TLKASYEKPRTGLVETKVKVYSNRRLSRLFGSKDITTTKFTPEYKDMANKFAPDG 1440
 DB 1381 TLKASYEKPRTGLVETKVKVYSNRRLSRLFGSKDITTTKFTPEYKDMANKFAPDG 1440
 QY 1441 SFARCYIDLOQFEDQITGKASQFDLNCFENETMNSNGNQPKRGKRYKIAOLEVMKLYVP 1500
 DB 1441 SFARCYIDLOQFEDQITGKASQFDLNCFENETMNSNGNQPKRGKRYKIAOLEVMKLYVP 1500
 QY 1501 RSDPRELIPTSIRSAVESINELNNEQNNYFEGYLHQEGGDCPIFKKRFKLMGTSILAS 1560
 DB 1501 RSDPRELIPTSIRSAVESINELNNEQNNYFEGYLHQEGGDCPIFKKRFKLMGTSILAS 1560
 QY 1561 EISHKTRAKINLSKVVDLIYDKENIDRSNHRNFSVDVLLDHAFTKIRANGELIDFCAPN 1620
 DB 1561 EISHKTRAKINLSKVVDLIYDKENIDRSNHRNFSVDVLLDHAFTKIRANGELIDFCAPN 1620
 QY 1621 KHEMKIWIQNLQETIYRNRFRROPVWNLMLQOQOQOQOQOQSSOQ 1664
 DB 1621 KHEMKIWIQNLQETIYRNRFRROPVWNLMLQOQOQOQOQOQSSOQ 1664

RESULT 2
 AAU79331
 ID AAU79331 standard; Protein: 1664 AA.
 XX AAU79331:
 DT 02-JUL-2002 (first entry)
 XX
 DE Candida albicans integrin-like protein, Intlp propeptide.
 XX
 KW Integrin-like protein; Intlp propeptide; bactericide; vaccine;
 KW yeast infection; thrush; paronychia; candidal vaginitis;
 KW immunocompromised host; cancer patient; transplant patient;
 KW premature newborn; human immunodeficiency virus infection;
 KW HIV infection; periodontitis; oral ulceration; esophagitis.
 XX
 OS Candida albicans.
 XX

Key Location/Qualifiers
FT Binding-site /label= Heparin_binding_site
FT 155..169
FT Domain /label= Catalytic_domain_1
FT 435..639
FT Domain /label= Catalytic_domain_2
FT 738..949
FT Domain /label= P-domain
FT 1022..1236
FT /note= "Processing domain"
XX WO200226257-A1.
XX 04-APR-2002.
XX 28-SEP-2001; 2001WO-US30312.
XX 28-SEP-2000; 2000US-237082P.
XX (UYA) UNIV YALE.
XX Hostetter MK, Devore-Carter D;
XX WPI; 2002-330056/36.
XX N-PSDB; ABK48292.
XX
XX Antibodies that react with the Intlp pro-peptide or peptide, useful for
XX PT vaccinating against infections by Candida albicans and Saccharomyces
XX cerevisiae .
XX
XX Disclosure; Fig 2A-B; 53pp; English.
XX
XX The invention describes an isolated antibody generated against the
XX pro-peptide of the Intlp protein of Candida albicans or a peptide region
XX of the Intlp protein of C. albicans involved in the activation of the
XX Intlp pro-peptide. The antibody may be administered for the treatment of
XX diseases caused by C. albicans and Saccharomyces cerevisiae (e.g. in the
XX form of a vaccine) and for detecting the presence of Intlp polypeptides
XX in samples for the diagnosis of these diseases. The antibody can be used
XX to treat C. albicans infections including thrush in newborn infants,
XX paronychia and candidal vaginitis. The antibody can also be used to treat
XX C. albicans infections in immunocompromised hosts such as cancer
XX patients, transplant patients, premature newborns or patients infected
XX with human immunodeficiency virus (HIV) on which C. albicans causes
XX diseases including periodontitis, oral ulceration and esophagitis. This
XX is the amino acid sequence of the C. albicans integrin-like protein,
XX Intlp propeptide that plays a major role in activating T-lymphocytes and
XX can be cleaved to release a superantigen-like moiety.
XX
XX Sequence 1664 AA;
SQ
Query Match 100.0%; Score 8631; DB 23; Length 1664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

241 OLLNNNEVSEPEALDTMKLRENFSLDEKVNLYLSPPTNNNNKSNVSDMSHLQNL 300
QY 301 OQASKKTKNTENTHNLISFALKAPKNDLENPLNSTJNNDISLRSSGSSQSLQSRNRPVL 360
Db 301 OQASKKTKNTENTHNLISFALKAPKNDLENPLNSTJNNDISLRSSGSSQSLQSRNRPVL 360
QY 361 ESWPGSPKKVNPGLSLNDIKGFSDEVESLLPRDISRKLFTYTKRHDPPEHNENFTIA 420
Db 361 ESWPGSPKKVNPGLSLNDIKGFSDEVESLLPRDISRKLFTYTKRHDPPEHNENFTIA 420
QY 421 KSTNTNKGQLLVSSDDHLSDFRSYNHTBOSILNLLNSASQSLSLNALEKQRTQEQEQ 480
Db 421 KSTNTNKGQLLVSSDDHLSDFRSYNHTBOSILNLLNSASQSLSLNALEKQRTQEQEQ 480
QY 481 TQAAPEEETSFSDNITKVQEKSNLEFYKVTIKKEPVATETKAKKREFSSILIKNE 540
Db 481 TQAAPEEETSFSDNITKVQEKSNLEFYKVTIKKEPVATETKAKKREFSSILIKNE 540
QY 541 DELAEPADITHPKKENANSHVEDTDALLKKALNDEESPTONSTKMSIRFHIDSMKLE 600
Db 541 DELAEPADITHPKKENANSHVEDTDALLKKALNDEESPTONSTKMSIRFHIDSMKLE 600
QY 601 DSNDDGREDNDDISREKSDILNDYSQTSIDIGDKYGSSSEITTKTLAPPRSDNDKEN 660
Db 601 DSNDDGREDNDDISREKSDILNDYSQTSIDIGDKYGSSSEITTKTLAPPRSDNDKEN 660
QY 661 SKSLEDPANNESIQQOLEVPHTKEDDSILANSSNINAPPELTIPLYEANDYSFNFVTKT 720
Db 661 SKSLEDPANNESIQQOLEVPHTKEDDSILANSSNINAPPELTIPLYEANDYSFNFVTKT 720
QY 721 FAYSSFEESLSREHETDSKPINFISIMHKQEKQKHQIKVPTKQIILASVOYKKNEOS 780
Db 721 FAYSSFEESLSREHETDSKPINFISIMHKQEKQKHQIKVPTKQIILASVOYKKNEOS 780
QY 781 RYTSQVKYIPNAIQEKKEEVNYSRRVYSPMDNLVNSOFLPELSBDSGFKLNPANYS 840
Db 781 RYTSQVKYIPNAIQEKKEEVNYSRRVYSPMDNLVNSOFLPELSBDSGFKLNPANYS 840
QY 841 NNTNRRSTPLSTKVLNINDPNVPEPEKSAETIRNARRLSANKAAPQAPPLRP 900
Db 841 NNTNRRSTPLSTKVLNINDPNVPEPEKSAETIRNARRLSANKAAPQAPPLRP 900
QY 901 QRPSTFESNSNKRYSRFPVPEIRTSALAPCOMYNDIFPDFGAGSKPTIKAGCMKT 960
Db 901 QRPSTFESNSNKRYSRFPVPEIRTSALAPCOMYNDIFPDFGAGSKPTIKAGCMKT 960
QY 961 LPSMDKDVKRIINAKKGYTQDEY INAKLVQDKPKKNSIVTDPEDRYEELQQTASIHMT 1020
Db 961 LPSMDKDVKRIINAKKGYTQDEY INAKLVQDKPKKNSIVTDPEDRYEELQQTASIHMT 1020
QY 1021 IDSSITGRPDSTIDMLPTLSDCLKRPPTALLSADRLFMQEVNHPRLNSVLVHPGAGA 1080
Db 1021 IDSSITGRPDSTIDMLPTLSDCLKRPPTALLSADRLFMQEVNHPRLNSVLVHPGAGA 1080
QY 1081 TNSMMLPEPDFELINSPARVNSNSNDVAISGNASTISFQDLDMNFDDQATIGOKIQEOP 1140
Db 1081 TNSMMLPEPDFELINSPARVNSNSNDVAISGNASTISFQDLDMNFDDQATIGOKIQEOP 1140
QY 1141 ASKSANTVAGDDDGLASABETPTPKKESISSKPAKLSSASBPRIKIGSPVRIKKN 1200
Db 1141 ASKSANTVAGDDDGLASABETPTPKKESISSKPAKLSSASBPRIKIGSPVRIKKN 1200
QY 1201 GSIAGIEPTPKATHKKKSFGQNEISNHKVROGDISPSSGSEHQNNPNSVSPSOYTDA 1260
Db 1201 GSIAGIEPTPKATHKKKSFGQNEISNHKVROGDISPSSGSEHQNNPNSVSPSOYTDA 1260
QY 1261 TSTVPDENKDVQKPREKQKKNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNH 1320
Db 1261 TSTVPDENKDVQKPREKQKKNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNHNNH 1320
QY 1321 VLGIKINILPDINTHKGRFTLLDNGVHCVTTPREYNNMDHNVAIGKEFELTVADSLFTL 1380
Db 1321 VLGIKINILPDINTHKGRFTLLDNGVHCVTTPREYNNMDHNVAIGKEFELTVADSLFTL 1380


```

QY 841 NNTNRPSFTPLSTKKNVLSINIDNDPNVVEPPPKSYAEIRNARLSANKAANOAPPLP 900
Db 841 NNTNRPSFTPLSTKKNVLSINIDNDPNVVEPPPKSYAEIRNARLSANKAANOAPPLP 900
QY 901 QROPSTSNSSNKKRVSRFRVPTFEIRRTSSALAPCDMYNDIFDDGAGSKPTIKAGCKMT 960
Db 901 QROPSTSNSSNKKRVSRFRVPTFEIRRTSSALAPCDMYNDIFDDGAGSKPTIKAGCKMT 960
QY 961 LPSMKDQVKRLNKKGVTOOEYINAKLYDOKPKKNSIVTPEDEPVEELOQTASTIHNT 1020
Db 961 LPSMKDQVKRLNKKGVTOOEYINAKLYDOKPKKNSIVTPEDEPVEELOQTASTIHNT 1020
QY 1021 IDSSLYGRPDSISTDMLPLSLDELKKRPALLSADRLFMEOEHPHRSNSVYLVHPGAGA 1080
Db 1021 IDSSLYGRPDSISTDMLPLSLDELKKRPALLSADRLFMEOEHPHRSNSVYLVHPGAGA 1080
QY 1081 TNSMILPEPDFELINSAPNVSNSNDVAISGNASTISFNOLDMNFDDGATTGOKIOEQP 1140
Db 1081 TNSMILPEPDFELINSAPNVSNSNDVAISGNASTISFNOLDMNFDDGATTGOKIOEQP 1140
QY 1141 ASKSANTVAGDDDLASAPETPRTPKESISSKPAKISSASPRKSPKIGSPVRIRKKN 1200
Db 1141 ASKSANTVAGDDDLASAPETPRTPKESISSKPAKISSASPRKSPKIGSPVRIRKKN 1200
QY 1201 GSIAIEPIPKATNPKKSFQNEISNHRVDRDGLSPSSGSEHOQHPMSVSPQYTD 1260
Db 1201 GSIAIEPIPKATNPKKSFQNEISNHRVDRDGLSPSSGSEHOQHPMSVSPQYTD 1260
QY 1261 TSTVDENKDVOKPREKOKNNHNRHHNHHKOKTDIPGVVDEIPDVLQERKLFPR 1320
Db 1261 TSTVDENKDVOKPREKOKNNHNRHHNHHKOKTDIPGVVDEIPDVLQERKLFPR 1320
QY 1321 VLGITNINLPDINTKGRFTLLDNGVCHVTTPREYMDNDHNAIGKEFEELTVADSLFTL 1380
Db 1321 VLGITNINLPDINTKGRFTLLDNGVCHVTTPREYMDNDHNAIGKEFEELTVADSLFTL 1380
QY 1381 TLKASAYERGTLVEYTEKKVYSRNRSLRSLFGSKDITTTKFPVPEVADTMANKFAPDG 1440
Db 1381 TLKASAYERGTLVEYTEKKVYSRNRSLRSLFGSKDITTTKFPVPEVADTMANKFAPDG 1440
QY 1441 SFARCYIDLOQFEDOTGKASQFDLNCFNEMWETMSNGNOPMKRGPKYKIAQLEVKMLYP 1500
Db 1441 SFARCYIDLOQFEDOTGKASQFDLNCFNEMWETMSNGNOPMKRGPKYKIAQLEVKMLYP 1500
QY 1501 RSDPREILPTSRSAVESINEINNEQNNYFEGYLHOEGGDCPIFKKRFKLMGTSLAHS 1560
Db 1501 RSDPREILPTSRSAVESINEINNEQNNYFEGYLHOEGGDCPIFKKRFKLMGTSLAHS 1560
QY 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVDLLDHAFTKFAANGELIDFCAPN 1620
Db 1561 EISHKTRAKINLSKVVDLIYVDKENIDRSNHRNFSVDLLDHAFTKFAANGELIDFCAPN 1620
QY 1621 KHEMKIWIQNOLEIIRNFRQPVNMLQOOOQOOOQSSQO 1664
Db 1621 KHEMKIWIQNOLEIIRNFRQPVNMLQOOOQOOOQSSQO 1664

```

RESULT 4

AAW99456
ID AAW99456 standard; protein; 236 AA.

AAW99456;

08-JUN-1999 (first entry)

Amino acids 218-453 of C.albicans integrin-like protein 1.

Integrin-like motif; vaccine; immune response; antibody; inhibition;
adhesion; endothelial cell; pathogenesis; infection.

Candida albicans.

```

PN US5886151-A.
XX 23-MAR-1999.
XX 03-MAY-1996; 96US-0642846.
XX 03-MAY-1996; 96US-0642846.
XX (MIND ) UNIV MINNESOTA.
XX Bendei CM, Gale CA, Hostetter MK, Kendrick K, Tao NJ;
XX WPI: 1999-242618/20.
XX New isolated Candida albicans protein with integrin-like motifs
XX Claim 1: Column 35; 21pp; English.
XX Peptides AAW99456-w99461 are derived from a Candida albicans protein
XX with integrin-like motifs, alpha-IMP1. This sequence represents amino
XX acids 218-453 of alpha-IMP1. The peptides can be used for producing
XX vaccines for stimulating an immune response. The antibodies can inhibit
XX the adhesion of C.albicans to cells, particularly endothelial cells.
XX This blocking activity of the adhesion to cells can reduce or prevent
XX subsequent events in the pathogenesis of invasive candidal infection.
XX Sequence 236 AA:
SQ
Query Match 14.0%; Score 1211; DB 20; Length 236;
Best Local Similarity 100.0%; Pred. No. 5.2e-67;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 SDEDTNNAVPPRPPLHTTPTFAOLLKKNNEVSEPEALTDMLKRENFSLSLDEKYNL 277
Db 1 SDEDTNNAVPPRPPLHTTPTFAOLLKKNNEVSEPEALTDMLKRENFSLSLDEKYNL 60
QY 278 YLSPPTNNNSKNVSDMDSHLQNDASKNKTENINHLSPALKAAPKNDIENPLNSLTNAD 337
Db 61 YLSPPTNNNSKNVSDMDSHLQNDASKNKTENINHLSPALKAAPKNDIENPLNSLTNAD 120
QY 338 ISLRSSGSSQSLQSTRNDNRVLESYVGPSPKKNVPGISLNDGIKGFSDEVESLPRDLS 397
Db 121 ISLRSSGSSQSLQSTRNDNRVLESYVGPSPKKNVPGISLNDGIKGFSDEVESLPRDLS 180
QY 398 RDKLETTKEHADAPENHNENFTDAKSTNTNGOLLVSSDOHLDSPDSYNTQESIL 453
Db 181 RDKLETTKEHADAPENHNENFTDAKSTNTNGOLLVSSDOHLDSPDSYNTQESIL 236

```

RESULT 5

AAE19800
ID AAE19800 standard; protein; 236 AA.

AAE19800;

18-JUN-2002 (first entry)

Candida albicans integrin-like protein (alphaIntlp) fragment.

Integrin-like protein; alphaIntlp; gene; fungicide; contraceptive;
HIV; human immunodeficiency virus; vaccine; cholera.

Candida albicans.

US6346411-B1.

12-FEB-2002.

08-MAR-1999; 99US-0264604.

03-MAY-1996; 96US-0642846.

(MIND) UNIV MINNESOTA.

Qy	530	FSSRILRIKKNEDDIAERADIPHRKKEVEA-NHVEEDTDALLKALNDEDEESTTONSKMS	568
Db	1209	---AVGPPDKTETSSSTNIIDKSSNESFDSMAQPSD-----RLNQKSAFTKLSSISSP	1260
Qy	589	IRPHIDSMKL-----EDSDGORENDIDISPREKDI-----LNDV	625
Db	1261	KKIMKODKLDLALSKGGSNPIIRDTGDSQOTQKKEQENDYKHEEDSKLANIDET	1320
Qy	626	SQTSIDIIGKYGNSSSEITTK-TLAPRSDNNKDESKSLED-----PANNESL	673
Db	1321	KSSSEKDAEPISKDSSQDSAKPRLSPKRGKKRKKNEKKPNDISIAESDIEGQVMTETV	1380
Qy	674	QOOLEP-ETKEDDSTLANSNAP-----PELTLPVVENDY	711
Db	1381	QACSTPSESNKDKMYSDETEBEPNLSTETIGRIKRGQAFHIEPNKD-DLHTTPQEN	1439
Qy	712	SSFNDDYKTPDAYSPSESLREHEDSKPINFISIMHKQKKHOIKVPTKOTIASY	771
Db	1440	QSTAGV--NEKQVPLPESV---ESDI-----PLMKLPTTYLMCT	1475
Qy	772	QOYKNEDESRVTSDKVKIPNALQFKKFEKVENMSRRVYSPMDIDLNSQFLPELSEDSGF	831
Db	1476	KNKSTLSSAEDPDYLER-----QXLITTSKGDSNPLDMMN-----NLEHSST	1520
Qy	832	KDLNFANYSNNTNRPSRFPPLSTKNVLSINDPNVVERPE-KSYAIEIRARLSANKA	890
Db	1521	ODPKEHEFSBOT-----FT--DMSDIIIPSCSTKSOIIVEPPTTPKSSDQTKN-----SFI	1567
Qy	891	APNQPPLP-----POQOPSSTSNKNSKVRNRVPTPE-----IRTS	929
Db	1568	TPNNSPSPKSRKNVSKAKRLDNSFEESONASSESSASKVOKELRTPYASCRLRYLIKTRP	1627
Qy	930	SALAPCMMYNDIEDDEGAGSKPTIK-AEGMKPLPSMD-----KDDYKRI	972
Db	1628	TSSLPTNRSKRSIKKTPAPKSKRLTLTLEMEXTPSRSPSVSLGEVNPDSQVVAESAVAL	1687
Qy	973	LNAKGYTODEYINAKLVODKPKKNSIVTD-----PEDRYE-ELQO--PASHINATIDSSI	1025
Db	1688	HESDRDLESNEIPENEVEFEDTEASAEPTDNKLRKKEDDHELEVUDICAAKNPITDDST	1747
Qy	1026	YGRDJSIT--DMLRPLSDLEKPKPALLSAD-----RLMPOEVPPLRS-----	1068
Db	1748	KDASNSKSTSDVLOETKIDL---NSLINAQGEDTPRIKELTEBEVNNKNTVEDESKQ	1804
Qy	1069	-----NSVLVHPGAGAAITS-----SMLEP-----DFELINSFAR	1099
Db	1805	EILKDEPRMAALEEDPTASAKAAEBMDLIIEKSNVSVYLAEPTDVTDDBELAQSPIR	1864
Qy	1100	NVSNNSDNVALISGNASTISF-----NOLDMNFDOAT-----IGQKIDEP	1140
Db	1865	N-SETTSVTDDEPSTSVYKRSLLKRRRADSSOPREDAKRRKODVEXSLKKEQVPR	1923
Qy	1141	A-----SKSANTVVG-----DDGLASAPERTPTPKKESISSK	1174
Db	1924	ARRQOLAEVERBLSKRSKTESFAKSTVOGKUIIITGNETIINSTTARPRETNREASTS	1993
Qy	1175	PAKLISA-----SPKRSPIKIGSPRYVLKXGSIAGIERIPKAT--HK	1215
Db	1984	PSAKSAVQEAKEKHNETTKHILDRPCKKLLHDSRPAEKK-----PMVOTLLSSTLSQK	2039
Qy	1216	PKKSFQGNESISNHRV-----DGLSPSSGSEHOONRPNMWSVSPSOYTDAATST	1263
Db	2040	PSLTDSDSPL--KIRKSLKSIADENIDODGSIFSSSVLKNKNTSVYA-PRKIVISVSL	2095
Qy	1264	VPDENKVO-----HKPRKQOKJNNNNNNNNNNNNNNKQOTDIRGVUDDER	1307
Db	2096	L--OSKTOVETLAASSETPILTKKEELKQOKTKPRBESKNKTESKKSLYOG-----P	2147
Qy	1308	DVGLQEGKLEFFRYLIGIKINP-LPDINTINGRFTLIDGVSUCVUTPREYNNMDINVALIG	1366
Db	2148	QMKQOKSE--AVSGKILNKLYLKSBTESKRTIVTATG-----RKQIOG	2190
Qy	1367	EFELTVASLEFIILTLKASYEKRGPTLVEYTEKKVUKSNRLSRLEGSKDIIITTKFVPT	1426

Db	2191	-----	LEVL--	KKRESKRSESLVEALSRKQSOYQRLSKIDGKSESTSLPQDV	2239
QY	1427	EVKDITWANKFAPDGSFARCYIDLOOFEDQITGKASQFDLNCFNEMETJNSNGNQPKRCKP	1486		
Db	2240	SKSEFALKAALP-----	-----	KETEFPPVQ--DALEIKMSKGGHNAVKN	2278
QY	1487	YKIAOLEVKKMLVPRSDPRELLPTSIKRSAY-ESINELANNEONN	1528		
Db	2279	TKTEQ-----	PKSKPK---	TEVRSLOAAATELMDSMS	2309
RESULT 7					
ABP38314	ABP38314	standard; protein; 10182 AA.			
XX	ABP38314:				
XX	AC				
XX	ABP38314:				
XX	24-JUL-2002	(first entry)			
DT					
XX	Staphylococcus epidermidis	ORF amino acid sequence SEQ ID NO:3159.			
DE					
XX	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;				
XX	antibacterial; gene therapy.				
KW	Staphylococcus epidermidis.				
XX					
OS	US6380370-B1.				
XX					
PN	US6380370-B1.				
PD	30-APR-2002.				
XX					
PF	13-AUG-1998;	98US-0134001.			
XX					
PR	14-AUG-1997;	97US-055779P.			
XX					
PR	08-NOV-1997;	97US-064964P.			
XX					
PA	(GENO-) GENOME THERAPEUTICS CORP.				
XX					
PI	Doucette-Stamm LA, Bush D;				
XX					
DR	WPI: 2002-381255/41.				
XX					
DR	N-PSDB: ABN90859.				
XX					
PT	Novel isolated nucleic acid encoding a Staphylococcus epidermidis				
XX	polypeptide, useful for diagnosing and treating bacterial infections -				
PS	Disclosure: SEQ ID 3159; 267pp; English.				
XX					
XX	ABN90358 to ABN93374 represent Staphylococcus epidermidis open reading				
CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences				
CC	given in ABP3124 to ABP37960. The S. epidermidis sequences have				
CC	antibacterial activity and can be used in gene therapy. The sequences				
CC	can also be used in the diagnosis and treatment of bacterial infections,				
CC	particularly S. epidermidis infections. The sequences can be used to				
CC	screen for compounds able to interfere with the S. epidermidis life				
CC	cycle or inhibit S. epidermidis infection.				
CC	N.B. The sequence data for this patent did not form part of the printed				
CC	specification, but was obtained in electronic format directly from the				
CC	USPTO web site.				
XX					
XX	Sequence	10182 AA;			
Query Match					
Best Local Similarity 17.9%; Pred. No.5,3e-13;					
Matches 363; Conservative 326; Mismatches 688; Indels 650; Gaps 91					
QY	34	KPLNPPRINSKPSLDPNSSD--TYSQDQKGEKKD-----TAFQTSFDRNFDD	85		
Db	3424	EPLFETAKNOLOANIDQKPTDGMTOOSVOSYERKLOEAKDKTINSINNVLANPDVAIRPT	3483		
QY	86	NSIDIQQTIOHQOQPOQ-----OQQLSTQDNNLIDEFSGFQIPMTSTDLTKRON-----	134		
Db	3484	NVEYEQ--INNELTQAKQGLTVQDKPLINAKTRALOOSLDNOSPSTGMTEATIQNIYNAKQ	3542		

[illegible]

OY	924	EIRRTSSSLAP-----	---CMVNDIPDRDGCAGSKPTIKAEGLTKLPS--MDKDYKR	971
Db	4541	KAQVMS---ARNNHITLETIKNAKADTLN---	OSMTALSESEIADYEKOKOQENTLDSNNKR	4594
OY	972	-----ILNAKKGV---	TODEYINAKLYDOKR---KNSIVTREDREYEELOQTASIHNAV	1020
Db	4595	QDYDAVMAAAGILNQTOSPTMSADYDQKAEEDVUKRTKLTALDGNORLEVAKQOALNHNHT		4654
OY	1021	IDSSIYGRPDSISTDMLRYLSEDELKKRPTLLSADRLFLMEQVEHPLRKSIVLYVHRCAGA	1080	
Db	4655	LND-----	---LNDARQGLTDTTL-----HSEPNINSVQAQEKANT	4688
OY	1081	TNNSM-----LPREDEL-----INS--PARVYNSNSDVAISGNASTISFQNDMNF	1127	
Db	4689	VNTAMTQKQJTIANYUDELDHNGVNIADKDKQKAYUNNAVN-----NAQL-INQSDANO	4742	
OY	1128	--DOATGQKIQEORPASKSANTYAGDD-----	---DGLASAPETRPRTTKESIS	1172
Db	4743	QLDRLEIKVQVQRVNTTK--NDLNGINDKLAEMARKDANTTIDGLTFLNEQARKKA--KEVNG	4799	
OY	1173	SKPALSSAPRSKPRKIGSGPVRYIKNGSIAIGERIPKATNPKKSFQNGEISNHHKYVD	1232	
Db	4800	KASTKTNTTSGLDQDYNDOLNIAHQALRS-----	---YNOVNVKANS	4838
OY	1233	GGISPSGSE-----	---HOQNPSPV--SVPSQVTDATSTV-----	1264
Db	4839	MYINDNDKPKVAYQNAVTHAQTLINQOSNPEMSRDYDVKQTOAVNTHANOHLGQOQLEQA	4898	
OY	1265	--PDEKDYONKPREKOKKHNNHNNHNNHNNKOTDIP-----	---GVYDDEIPVIG	1310
Db	4899	QSSANTELGNIPLNTNQKAKEKELVNSKQTRIEVDQOLQAOKSLDSSQGTAKSLVAKOR	4958	
OY	1311	LOERKGLFEFRLVIGIKINILNDPINTHKRGFTL-----	---TLDNGVHCVTTPRE	1354
Db	4959	TVQKTSVYI-----NEDQEOSAYUNDSTMQOTIINKTADPVLDKTLDVNAISMTKE	5012	
OY	1355	YNMDDHNAIIGKEBELVVA--DSLFEPLILKASYEKRPGTIVEV-----	---TEKKYVKSBN	1406
Db	5013	N-----ALHGEBKLTATETAINALNTPL--ALNTPQOEAKIKTALINTANTHTDVTAAOS	5064	
OY	1407	RLSLRFGS---KDLITTFKRVLEVUKDTWANKFARDGSPARCUID--LQOFEDQITGKAS	1461	
Db	5065	KANOINSAMHTLRQNISDSNYSVNE-----SNVYNAREKONAFTEALNNAKEIYNEGOA	5119	
OY	1462	QFDLNCPE-----WETNSNGNORPKCKPKTKIAQLEVKMLYVPRSPREILFTSTIRS	1514	
Db	5120	TLDAINSIQKAOALITLTKNALDDEQRLRRK-----	---ENADQEI-----	5157
OY	1515	AYESINELNPNQNNYFEGYULHOEGDQDPRK-----KREFKLMGTSLLASHESISKTRAK	1569	
Db	5158	--NLNLQTLDRQRSEGLINSQSTRFEVASQALAKKAKELNKYA-----EDLNHLINGK	5208	
OY	1570	---INLSKVVDLIYDKENIDRSNHRNFSVDLLLDHAFRIKFRANGEL	1613	
Db	5209	NOMINSSKFI-----NEDANQOQAYSAIASAALKNKKSQNPBL	5247	
RESULT: 8				
ID	ABP40235	standard; Protein; 3696 AA.		
AC	ABP40235;			
XX	DT	24-JUL-2002 (first entry)		
DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.			
XX	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;			
KW	antibacterial; gene therapy.			
XX	Staphylococcus epidermidis.			
XX	US6380370-B1.			

OY 1306 IPDVG--LOE-----RGK-----LEFRVLGIKINLP-----DINTH 1335
 DB 2661 RQESAKQLQDEVLKKAIAIKDGQITNDVEKTVVNGIAETENI-LPATTVKDKAKADVNAE 2719
 OY 1336 KGRFTLLTDNGVHCVTTEPEYNMDDHNAVIGKEFELTVADSLFPL--TLKASYEKPGTL 1393
 DB 2720 RQKQNLQINSNDQAT-----EKLVSADNLNHHVETTNQALIDAPPTNQ 2764
 OY 1394 VEYTEKK-----VKSRRNLSLFGS-----KDIITTKRYPTVEYKDTWANKFA 1437
 DB 2765 VNVEKNKIGTIRDIQPLVVKPKTKSIEEVAVEKKKTEIQTQONATQHDVEVRE-GLNQLN 2823
 OY 1438 PDGSPARCVIDLQOEFDITGKASQFDLCNCF---EWEETSMCNQNPARKRKPKYKINOLE 1493
 DB 2824 QIHEKAKMDVMSQTNQOYE--NAEONSIDQITNNEPDESKSRMNAVEALVKAQOKKIDETE 2882
 OY 1494 VKM--LYVPRSDPREILPTSIKSAVESINELN--NEONNFEGLHDEGDCPIFKKRF 1549
 DB 2883 QEFSATQREKDNALQHDQYKEIINSINQANTQNEVDNATSGLNITTEPRPYMKK-- 2940
 OY 1550 KIMGTSLLAHSEISIKTRAKTNLSKVVDLIYDKENIDRSHNRNFSVDLLDHAFTKFA 1609
 DB 2941 --KKAAILKLVDSPTQEAIIIN-----GYPDATDELDQANSKLKILDDAKKQGLA 2990
 OY 1610 --NGELIDFCAPNKHHEM 1625
 DB 2991 HTNNEVDIDYNEVSQMK 3008
 RESULT 9
 AAG82935
 ID AAG82935 standard; Protein: 5024 AA.
 AC AAG82935;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:2964.
 XX
 KW Staphylococcus epidermidis sri strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 PD 17-MAY-2001.
 XX
 PE 09-NOV-2000; 2000MO-USJ0782.
 XX
 PA 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PT kimmerly MJ;
 DR WPI: 2001-316495/33.
 DR N-PSDB: AAH53785.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 18; Page 779-781; 2188bp; English.
 CC
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81455 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. endocottidits. AAH5391 to
CC AAH5090 represent specifically claimed *S. epidermidis* genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to
CC AAH5098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472
CC no sequences are present for SEQ ID NO:4455 to 4464.

SQ Sequence 5024 AA;

Query match	Score	DB	Length
3.98;	338;	22;	5024;

Matches 388; Conservative 309; Mismatches 714; Indels 560; Gaps 91;

QY 24 SSASIFNSPTKPLNFRT-----NSKPSLDPNSSSDTYTSEQDQEKGEKKDTAFQ 75

Db 772 SSAYINADPTKVTA YDQALQNAENI NATPNVELNKA-----TIEQALSRVQQAQ----- 822

QY 76 TSFDRNFDLNSIDIQTIQHQQQQPQQQQLSOTDNNLIDEFSSQTPTMTSTLDLTQN 135

Db 823 -----DLDG---VQOLANAKQQTQTVNGLSLNDGQKRELLINSANT--RTKVQE 870

QY 136 TVDKVNE-NHAPTYI-NTSPNKSIMKKATPKASPKVAFTVTNP EIIHHYPDNREEEDQS 193

Db 871. ELNKATESNHAMEALRNSVQNVQVKQSNYYNEDQ-----PEQHNY-DNAVNEAQAT 922

QY 194 QQKEDSVEPLIQHWKDPQSQFNYSDEDTNASVPTPTPLHTTKPTF--AQLL----- 243

Db 923 --INNNAQPVL-----DKLAIERLTQTVNTTKDALHGTQKLIQDQQAEE 964

QY 244 -----NKNNNEVNSEPEALTDMKLRNFNSLSL-EKVNLYLSPTNNNSKNV 290

Db 965 TGIRGLTSLNEPQKNAEVAKVTAATTRDEVNRIRQEAATTLDTAMGLRKS IKDKNDTKNS 102

QY 291 SDMDSHLQNDASKNKTNENIHNLSPALKAPKNDIENPL-NSLTNADISLRSSGSSQSS 349

Db . 1025 SKYINEDHDQQAYDNAVNNAQHVIDETQATLSSDTINQLANAVTQAKSNLHGDTKLQHD 108

```

QY      350 LQSLRNDNRVLESVPGSPKKVNPGLSLN-----DCIKGFSDEVE--S 390

```

Db 1085 KDSAKQ̇TIAQLQ̇NLNSAQ̇KHMEDSLIDNESTRQ̇VQ̇HDLTEAQ̇ALDGLMGALKEIKDNT 114

```

QY      391 LIPRDLSRDKLETTYKE--HDAPEHNNENFIDAKSTNT--NKG-----QLLVSSDDHLDLS- 440

```

Db 1145 NIVSNGNYINAEPSKKQAYDAAVQNAQNIINGTNQPTINKGNVTATQTQVKNTKDALDGD 120

QY 441 --FDRSYNHTEQSILNL--LNSASQSQ-----ISLNALEK-QROTQEQEQTQAAPPEEET 490

Db 1205 HRLEEAKNNANQ TIRNLNLNNAQKDAEKNLVNSASTLEQVQQLQTAQQLDNAMGELRQ 126

QY 491 SFS DN I K V K Q E P K S N L E F V K V T I K K E P V S A T E I K A P K R E F S S R I L R I K N E D E I A E -- P A D 548

Db 1265 SIANKQVKADSK-----YLNEP-----QIKQNYDDAVQRVETIINETQNPPELLKAN 131

QY 549 IHPKKENEA^{NSHVED}TDAL^{LKKA}--LN^{DDEESD}TQ^{NSTKMS}IR^{FHID}SD^{WKL}ED^{SND}GD 606

Db 1313 I-----DQATQSVQNAEQALHGAELNQDKQTSSTE-----LDG---LTDLTDQ 135

```

QY      607 RE-----DNDDI-SRFEKSDILND-VSQTSDIIGDKYG-NSSEITTKTLAPPRS 653

```

Db 1355 REKLREQINTSNSRDDIKQKIEQAKALNDAMKKLKEQVAQKDCVHANSDYTNEDSAQKDA 141

QY 654 DNNDKENSKELEDPANNES-----LQQQLEVPHTKEDD-----SILA 690

Db 1415 YNN--ALKQAEDIINNSSNPNLNAQDITNALNNIKQAQDNLHGAQKLQQDKNTTNGAIG 147

QY 691 NSSNAPPEELTPVEANDYSSSENDVTKTF-----DAYSSFEESLSREHE-TDSKPI 742

Db 1472 NLNHLNQPQKDAL-IQAINGATSRDQVAEKLKEAEALDEAMKQLEDQVNDQDQISNSSP- 152


```

Db 2917 EEEOKHHPPEKKSSKSSSEOPKQSTEOYEISVTEHDLKP-----EERKPTVOYIOSET 2972
Qy 248 EV-----NSEEPA-LTPMKLRKRENFSL 270
Db 2973 NVEETKDDTGVHKOYTTKRLRRPAGELEIIEIVARDDOPEALITVIEPE---FVN 3029
Qy 271 LDEKVNLYLSPNTNNNSKNVSDMDSLONLODASKNKTN-ENIHLFSLAKPKNDEINP 329
Db 3030 ODEKPK---EPKKKTRKRYKKDDIDHYIOKLTIELETPKTELEKYEKEIEPEPVKOKPLDSP 3086
Qy 330 LNSLTNADISLRSSGSSOSSLSLRLND-----NRVLESYPGSP-----367
Db 3087 IDVLDESPEKQVKOKKSRSTKVPNEETPVQEQYAKVNVVEEAEPEQPEIIVQILEVYKV 3146
Qy 368 -----KVNPGSL-----NDGJKGSDEVESESL 392
Db 3147 EVDVKEVTEDEGKRVQEKTKRYLAKTIGTEBQTFKTIIMISEDNDSYTVIVDEPEELAS 3206
Qy 393 PROLSRDKLETTKEHADPEHNENFIDAKSTYNNKGQLVSSDDHLSFDRSYNHTEOSI 452
Db 3207 POSI-----EEH--PEOSKEKRLAPKPKTVRK---VKDD-----L 3237
Qy 453 LNLNSASOSQISLNALEKORQOTQOEOTQAAPEEETSFSDNKKVQEPKSNLEFVKVT 512
Db 3238 SDIYKKLIEETPKVVDLEKTEKE-----MEKPYKLTVSDSIEPEKRPD-----3282
Qy 513 IKKEPVGA-TEIKAPKREFSSRLIRKNE-----ETAEPAD 548
Db 3283 -KSQPIVLPTDTKPKTKPTKPTEDTQOVDPDEPTTVDTTIDPELPTQAOED 3341
Qy 549 -----HPKKENANSVIEDTALLKALNDESDTTONSTKSIHHDSDMKLEDN 603
Db 3342 TATQAITPSAQEESKSTQ-DTKRDTIOKTVKHKTKRPDOKSVETSELEPVHAKDOISIIH 3400
Qy 604 DG--DREDNDIOISFEKSDILNDYSOTSDI---IGDKYGNSSSETTKTLAPRSDNNDK 658
Db 3401 EELVEEQPEKILEVRYIDEVAEVESQPIVEEVEDEPPQATEVYEDVKPKS-----3455
Qy 659 ENSKSLDPAANNESLQOOLLEVPHTKEDDSILANSSNIAPEELTLPVVEANDYSSFNVDY 718
Db 3456 -----KKKRVKKKTDH-----DELKKMLE-----GEIE 3481
Qy 719 KT-FDAVSSFEESISREHEDSKRINFISTHKOEOKQKHQIHKVPTQOIIASVOQKNE 777
Db 3482 KTELEKTEKIEFDVPKKLEPEFALEPIKIERKKQ-----PKVYTL-----3524
Qy 778 QESRVTSDKVIAPAIQFKKFK-----EYNV-----MSRNVSPMODLVNSOFLPE 824
Db 3525 -----DATDVPTKVALKSKRKEKPAEELTVOLPKFRLKARV-----LVETPPAPLIPK 3574
Qy 825 LSEDSGFKDLNFANYSNNTNRP--SFTPLSTKVLNSIDNDPNVVEPEPKSYAEIRN 881
Db 3575 TTDTGAIKD--NGELSHNIEAEELIKFKPHKTKKI--KKIDOLEKVELEKYEKY-----3626
Qy 882 ARRLSANKAPNQAAPLIPQOROPSTSRNSNKKRVRVPFELIRRSSALAPCDMINDI 941
Db 3627 ---ISSEEPBEKTPYKKPEKAP-----PEEKEDV 3655
Qy 942 FDDPGAGSKPTIKAG-----MKTLPMSMDKDVRIINAKKGVTODEYINAKLVOKPKK 996
Db 3656 KIKGKGGKKKPEEAEENVTLKNIPOKPO-EVEEVELKOKPREVELEVEQO--TKRPKD 3712
Qy 997 NSIYTPDEKREYELQQTASIHMATIDSSIYGRPDSISTDM---LPLVLSDELKRP-PTAL 1051
Db 3713 GEFVVEP-----FEPSEFDRPEYVDELQIEHPLEPKVKKPKSKTKY 3755
Qy 1052 LSNARLEMEQEVHPLRNSVYLHPGAGATNSSLPPDEPELINSPARNSNNDNVAIS 1111
Db 3756 KPKOKSKSEPE-----TIVSELVAGVPKKEEAIIPQDVK-FRKPDEDAPEEDT-----3802
Qy 1112 GNASTISFNOL-----DMNFDQATIGOKIOEOPASKSANTVRDDDGSLASAPETPRPT 1166

```

```

Db 3803 ---SEIKLRPVQASKDEPNDEQALVTPK-AEERIPQETEDKALIDEKKPKSKPKVQOP 3858
Qy 1167 KKESSISKPAKLSSASPKK-----SPIKIGSP--VRVIKNGSIGAGIEPIPKATNPKPK 1218
Db 3859 KEQETAKEPEPEFEVSVKEEELNDKPIETIEKPKVKKKKEKPKAPVSEVVVVEEPEK 3918
Qy 1219 SFQGNELIS-NHKVNRDGGISPSGSEHOQNPMSVSPSQYTDATSTVDEKNDVOHKPRE 1277
Db 3919 EEPREIPEVEYKITTVLEPEDAP--KEHQVKVIDFDERQETTEEVI--EEKVYTRKKKP 3974
Qy 1278 KOKKHNNHHNNHNNKQTDPIGVYDDEIPVGOLOERKLEFRVLGGININLPDINTHKG 1337
Db 3975 KPQPEEFEEVTLKEPKEQIOPDVVASAI-SLPIEE-----PEQKPYQ 4017
Qy 1338 RETLTLDNGVHCVTTPPEXMDHNVAIGKEFELTVADSL--EFILTKASYEKP--RGTL 1393
Db 4018 EVELKITQ-----TTPF-EPNDVQIAYAEKTKRPAVKVYKEDKIYVVAEERKQVEETI 4071
Qy 1394 VEYTERKVVYKSRNLSRLFSKDIITTTKPV--PTEVKDTWANKFAPDGSFARCYIDLQ 1450
Db 4072 VEV-EKOEKKKSEKPKSYEER--ISETQISEKPIEVAEE-----APEET-----PK 4116
Qy 1451 QFEOQITGK--ASQFDLNCFMETMMSGNCPMK-----RGKPYKIAQLVYKMLYVP 1500
Db 4117 VEEKVAKKEPDSYEETLKEETDEKVTVDQPEEAPVAVVFKKPKPEPAVEAEFV---4173
Qy 1501 RSDPREILPTISRSA 1515
Db 4174 MTEPRIVETSVEYA 4188

RESULT 11
ABB60327
ID ABB60327 standard; Protein; 3111 AA.
XX
AC ABB60327;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7773.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW.
XX
DR WPI: 2001-656860/75.
XX
DR N-PSDB: ABL04430.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
PS Disclosure; SEQ ID NO 7773; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention

```


XX (ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
DR N-PSDB: AAS55262.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 12996; 511bp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from Wipo at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6281 AA:

Query Match 3 8%; Score 327; DB 22; Length 6281;
Best Local Similarity 17.7%; Pred. No. 2.5e-10;
Matches 296; Conservative 279; Mismatches 607; Indels 488; Gaps 71;

42 NSKPSIDPSSSDPTVSEQ--DOEGK-----KEKKK-----71
2459 NSGPNVITTAADVTATGORNNAETGINSPTNLATAKQKQKALRQMTLSDAKOSITG 2518
72 -----TAFQTSFDRNEDLNSID-IDQOTIQHQOQOPOQOQOOLSOTDNLIDFSPQT 122
2519 QIDSATQVGVGVQSVKDNATWLDNAMNQLRNSINKDKVAKSPYV--DADRRKQNAVNT 2575
123 PMNST---IDLTKONPTVQVKNENHAPTYINTSPKNSIMKATPKRASPKVATVTPPEI 179
2576 AVTMAENIINATISO--PTLDSAVTOAANOVST--NKTALNGAONLAN--KKOETTANINOL 2631
180 HHYRDNKREVED-----OSQKEDSEVERPLIOHWKD--PSOFNYS 218
2632 SHL--NNAKQKODLTOVTANPNISTVNOYKTAEOLOQAMERLINGLODKDYKOSINTF 2689
219 DEDTNAVSPPTPLHTTKPTFAQ--LINKNEVNSEPE---ALTDKCLKRENFNSLSLDE 273
2690 DAD-----DEKOTAYNNAAVTAAENIINQANGTNAQNSOYEAALSTVTTTKO---ALNGDR 2741
274 KYVLYLSPTNNNSKVVSDMSHLQNLQDASKKKTNTNHNLSFALKARNDLENPLNSL 333
2742 KY-----TDKNNANQTLSTLD-----NLNNAQKGAVTGNI-----NOAHVTV 2778
334 TNADISLRSSGSSSSLOSIRNDNRVLESVPGSPKKVNPGLSLND-----GIGFSDEVY 388
2779 AEVTOAIOTAOELINTAMGNLKN-----SLNDKOTTLIGSONFADA-- 2817
389 ESLLPRLDSKDLKLETKEDAPENHNENFIDAKSTNTN--KGOL-----LVSS 434
2818 -----DEKKNAVNEAVHNAENILN--KSTGTVNPKQDVEAAMQNVNTKAAING 2865
435 DHHLDSFDRSYNHTEOSIINLNSASQS-----QISLNALEKOROTOBEOQTOAEPREE 489

DB 2866 TQNLKAKQAHNTAIDGLSHLTNAQKEALKQOLVOOSTTVAEAGNENQKANNVDAADKLR 2925
QY 490 TSPSDNITKVKOEKSNLEFVKVITKKEPVSAITEIKAPKREFSRILIRIN--EDELEPAD 548
DB 2926 QSIADNATTKO-----NONYTDASQNK-----KQAVNAVTTAGCIIDOTTSPF-- 2969
QY 549 IHPKKEANEANSHVEDTALLKRALNDEESDTQNSTKMSIREPHIDSKWLEDSNDGDE 608
DB 2970 LDPTVINQAAGVSTT-----KNALNGENENLEAKQOASQSL-----GSLDLNNAQKO 3018
QY 609 D--NDDISREKSDIINDVQTSIDIIDKYGNSSETTTTL-----APPRSN 655
DB 3019 TVTQDINGAHVDEANQIKQMAQNLNTAGNLKQALADDAFKATVNTDADQAKQAVN 3078
QY 656 NDKENKSLDEPAN-----NESIQOOLEVPHTKEDDSILANS-- 693
DB 3079 TATVNAENIISKANGNATQAEVEQAIKQVNAKQALNNGNANVQHAKKDEATALINSNDL 3138
QY 694 NIAPREELTPVEANDYSEFNDYTKTF---DAYSFEBSLRHEHTDSKPINFISWH 749
DB 3139 NQAKQDALKQOVONATTVAGVNVKOTAOELNNAMTQLKGIADKQOT----- 3186
QY 750 KQKQKQHQIHKKVPTKOIIASVQYKNEQESRVTSPKVIIPNAIQKFKREVNVMSRVV 809
DB 3187 ---KADGNFVNADPDKQMAVNAVAEALISATPVDVVTPEIT-----TALMKV 3234
QY 810 SPDMDDLNVNQLPELSEDSGFKDLNFANVSNNTNBRPSTPLST--KNVLSNIDNDPV 867
DB 3235 TQAKNDNLN-----GNTN-----LATKQVNVQAHIDQLPVL 3264
QY 868 VERPEPKSYAETINARLSANKAPQAPRPPROPSPSTRSNKRVSRFPYTFEIR 927
DB 3265 NQAKRDEYSKQITQV-----TLVNVVNAIQQAATTLNDAMTQLKGIANKKQIIG 3314
QY 928 TSSALAPCDWYNDI--FDDFGAGSKPTIKAGM---KTLSPMDQDVKRLIAKKGYTQDE 983
DB 3315 SEN-----YHDAIDTKQATVDNAVTKAEELKQTTNPMTQNTTQALTKYVNTQNL 3367
QY 984 YINAKLVQK-----PKNSIVTPDRREELQOTASIHATIDSISYGR 1028
DB 3368 NGNQKLADAKQDAKTTGLTDLINDAQKQALT---QYQADADIATVNVNQKQANL-- 3421
QY 1029 PDSISTDMRLYSLDELKRRPTALLSADRLMEDEYVRLASVSLVYRGAAGATNSSMLPE 1088
DB 3422 ---NNAMTNLNNALODKTETLNSINFDDADKADATVATVASHAEGILSKANGSMAQO 3476
QY 1089 PDPE-----LINSPARVWNSNDVVALSGNAS--TISENODMNFDDQATIGQKIQEQPASK 1143
DB 3477 TEVEQAMQRYNE--AKQALGNDVNOVAKDAKQVITTNANDLNOQDALKQVDAQVYA 3535
QY 1144 SANTVGGDDGLASARETPRTPK--KESISSKPAKLSSAPRSPKISGPRVVRVIRKNS 1202
DB 3536 NVMTIK-----QTADDLNOAMTQLKQIADKQO-----TKANCN 3569
QY 1203 IAGIERIPKATPKRKSFGQNEISNKKVDRDGTSPSSGSEHQOANPSWVSQYTDATS 1262
DB 3570 FVNAD-----TDKQNAVNAVA-----HABQII-----S 3593
QY 1263 TVRDEKNDYONKPREKQ--QKNNHNNNNNNKQKTDIGVUNDELIDVGLDERGKLFPR 1320
DB 3594 GTPNAVNDVFOVAQALQOVNOQAKGLNGMNLQAVAKONANTADQLP----- 3640
QY 1321 VLGIKINILRDIPTNKGRETF--LTLDGVCVUTPREKMDHNVAIKFEELGVADSLFET 1379
DB 3641 ---NLNORQKALKDQVSHAEVYGVAAI---KQADALNNAMG----- 3678
QY 1380 LTLK-----ASYEKRGSTVEVTEKKVVKSR-----NRLSRLEGSKDITTTTKVPEVVK 1429
DB 3679 -TLKQOIQANNSQVPS--VDETFQADODKQOAVNNANQAOQIANG-----IPTPV- 3725
QY 1430 DTWANKFARDGSAFRCYIDLOQFEQDITG--KASQDILACFNEMETMSNGNQPMKKGKPY 1487

Db 3726 -----LTFD-TVTAQVTTMNAQKDALNGDEKIAQKQALANLDLRLDNQFORDALRN 3778
QY 1488 KIAQLEVKMLYVPRSDPRELPTISAVESINELNNEQNNPEGYHOE 1537
Db 3779 QINQAOAL-----ATVEOTKONAOVNMTAMSNLKQGIANKD 3814

RESULT 13
ABB68397
ID ABB68397 standard; Protein; 2768 AA.
XX ABB68397;
AC ABB68397;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31983..
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EM;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL12500.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2768 AA:
SQ

Query Match 3.8%; Score 325.5; DB 22; Length 2768;
Best Local Similarity 18.4%; Pred. No. 9,7e-11;
Matches 288; Conservative 244; Mismatches 646; Indels 383; Gaps 69;

QY 5 PSLULIDKSHIQLOPOSSASIFNSPTKPLNFPRTNKSPSIDPNSSSTIYSEDOEK 64
Db 425 PFO-IPEKETTEDELKIVSAPAKASPEEEV-VTATTSAPTEEDVKTPTAGTISEEEE 482
QY 65 GK-----EKKDPAQTSFDRNFDDLNDSIDIOOTLOHQOQPOOQOQLSOTDNLI 115
Db 483 GKTPAEBSGEEKVKYTAAPETEDEKAPTSAPVASEKQDEKRPESGDEE---L 539
QY 116 DEFSQTPMTSTLDLTKQNFVDKVNENHAPYIN---TSPNKS-----IMKA 161
Db 540 DLKPTAPTAGATSASEESEDEGKSTAPTSVDIDIEPAKPTESSEASGEGDVAKET 599

QY 162 TPKASPKKVA-----FTVNPDEIHHPDNKVEEEDSOQKEDSVEPPLIOHOMKDPQFN 216
Db 600 TP-AGEASIGEEIEIVKGTTP-----AGEPSSEGEIEIVKTPPAESESSESE-- 646
QY 217 YSDEDTNAPVPPTPLHTTKPTFAQLNKNNEVSEEPALDMLKLR-----ENSNLSLD 272
Db 647 --DELFTVTTTPAGPSPVAGEEELA-----KETTPAGEPSIAGEEIVAVTTTPAGSSIAGE 700
QY 273 EKVNLVLSPTNNNSKNVSDM-----DSHLQNLQDASKNKT-----NENIHL 315
Db 701 EEIVKVTTPAGESSESEGEIEIKVTTTPAGESSEGEIEIVKESPTPAGEPISEGEDEV--I 758
QY 316 SFALKAPKNDIENLNSLTNANDISLRSSGSSQSLQSLRDNKRLVSEVPSPKVNGLS 375
Db 759 KATTSAPKSDIEGVKEPETEVEPAEVEDFAKPTPIAEAE--EPITAGPIIP----- 810
QY 376 LNDGKGFSDVEVSSLPRDL-----SRDKLETTKEHDAREHNNENFIDA 420
Db 811 -TDGIGS-EEIEIVKGTTPQTLERQPEISEESTVEPVAEEDLSSTASALASTEGVQDA 868
QY 421 KSTNTKQGLLVSSDHL-----DSFDRSYNHPQSLINI-LNSASQSLSLNALERQD 474
Db 869 ASETTSAPARAGDKDEAATTVPTAODKDEEVEDATDLPVEDVQSTAKTTTTE--Q 926
QY 475 TORGEQTOAAPEEETSFSDINIKVQKPKS-----NLEFYK-----VITKKE--PVS 519
Db 927 PKESSTEADEALEVTSSPADKQVEPAEPADKDKDEEDVOTATDLPKSDIGPPV 986
QY 520 ATEI-----KAP-----KREFSRLIRIKN-----BDEI 543
Db 987 DTEATITGQPETSDETATDKPPSVYLLPVSQEVPESTAKVNDNRNFETEKPTLPESGEDQS 1046
QY 544 AEP-----ADHPKKEANEANSHVEDTA-----LLKALNDDDESQTTQ 582
Db 1047 SEPLPAMDLPAGIPGEGDCVLEGKTYANNTIVPATACDVSCSKCISLVAQOQCKEKLPE 1106
QY 583 NSTKMSIRFI-----DSDMKLEDSN-DGDEEDNDISRFESDILNVSQ 627
Db 1107 NLEKCTVADLDBCCPTTYICDESTESAERKDESTAKPDNKTIDVSEISTEELPKDVI 1166
QY 628 TSDIIGD--RYGSSSEITTKLAPPSSDNND--KENSKSLEDPAN----- 669
Db 1167 PTGITGQPLSHVKRDEDIQVTVSVAQFDESTAKVKKPIDESAEDKKPIGESEDSKP 1226
QY 670 -NESLQOQLEVPHTKEDDSILANSSNIAPPEBLTLPVVEANDVSSFNDVTKT--PDAYS 725
Db 1227 IDESEEDKKPVEESAEDKKPVEDESEKKEPLPVIPASEIEKESKPEDEKTEADFAAPT 1286
QY 726 SFERSL-----SREHETDSK-----PINFISIMHKQEKQKHQJHKNVTKQIATSYQ 772
Db 1287 EOPATTPOAIADPAEAEVDDKLAATTSAPVSGEDELKPADKKETEAQIIPDAIIPAS-- 1344
QY 773 QYKNEDSRVTSQDKIPNIAIQFKFKREVNVMSRRVSPDMDLNVSQFLPELSEDSGFK 832
Db 1345 --TDEPES-----STELPTVLDLKKPBEEDS--TKGTAPESDK-----VPEVPTS---- 1385
QY 833 DLNFANTSNNTNRPSTPLSTKNVLSINIDNPVAVPEPPEKSYAEIIRNARLSANKAP 892
Db 1386 ----ASTENEIEESDKFTTVAAPKISASDETEPAEEDLVATREPIESEPREVSTKRPAY 1441
QY 893 NQAPPLP-----PQRPSSSTRMSNKRVSFRFPTFEIRTSALAPCMKYNIFPDFG 946
Db 1442 -QGPPLPTLAPQEKRPVPAETSTEADIS--TEPSAEVEKEASG-ETSESDNI--DAG 1495
QY 947 AGSKPT- IKAEGMKTTLSSMDKDVKRIINAKKGVTQOEYIYAKIVDQPKKNSIVTQPED 1005
Db 1496 ASSTPVPVSADEDKT- PSTE-----KTYVADDKFTTVAALADDEESNIPKLPQD 1544
QY 1006 RYIELQOTASIHNAITIDSSIIYGRPDISITDMLPY-----LSDELKRPPTALASADRLFM 1059
Db 1545 IFEEDAPVAV--TTAASKDDGEGKPVAVEERKPIEDGQKPIEDETSTPTSS----- 1593

QY 378 DGKGFSEV---VESLP-----RDLRDKLETTK----- 405
| : : : : :
Db 859 VDAKGIONLIGHLTAKEVYPOTKLSPSTKEASALPKLVKTRPVLKSRSTVKGOK 918
QY 406 -EHDAPEN-----NNENFIDAKSTNTKQGLVSSDH----- 437
| : : : : :
Db 919 NPSEKPVHEEFLHPVSAVAGNRIPTSSISDDGSIFLGFDDNEGKRSKTPVRSQRYKOKM 978
QY 438 -----LDSFDRSYNHTEOSILNL-----LNSASOSQISLNALEKQROTQEOQTQAAE 485
| : : : : :
Db 979 INETISDADATPDYDETEEEQPSIEIQOPVKNSAADSKTFPDRHOVYGAFSDSDSSDNG 1038
QY 486 PEETSFSQDNKVKOEPSNLE---FVKYTIKKEPVASATEIKAPKREFS-----SRI--- 534
| : : : : :
Db 1039 PSSPMDFGDNMSIEDSRTSEEDVTAMAKLVNENEKPFVDEVLPSPKSGTEATVDCSIDD 1098
QY 535 -LRINDEDEIAPADIHPKKEANEASHVEDTDLKALNDEDESTTONSTKMSIRFIH 593
| : : : : :
Db 1099 PLKTGGEATSSHSEITPCASSEKCVTDE---KQPIKKEKSDITLDRKSEKSRMST 1154
QY 594 DSDMKLESDNDGRDNDNDISFEKSDILNDVSDIIGDKYGNSSSEITTKTLAPRS 653
| : : : : :
Db 1155 KSLMTTDEH-----VKDSSTTSKI-----NEKRLTSSPLKTPGS 1189
QY 654 DNN-----DKENSKSLEDPA---NNESLQOOLEV----- 679
| : : : : :
Db 1190 SENTELRLLEGNASDKQNEHLISPEALTNQKNALKNLEEVSPKPYSLRSRGRKS 1249
QY 680 -PHTKEDDSILANSSNIA---PPEELTPVYANDYSSFNQDYKTFDAVSSFEESLSREH 735
| : : : : :
Db 1250 GSHNKKLDESEKHSNISCACKEEKATERLEDGEOLTYVDSTE----- 1293
QY 736 ETDG---KPIFISIMHKOQKQKHQIHKVPTKQIIA---SYOQYKNEQESRYTSDKVKIPN 791
| : : : : :
Db 1294 DTDSTRNPIDNTAENTAIKRAK-----KTRCIYKSKOSKEMKDLKEDGVAKSKDDL- 1346
QY 792 AIOFKKFEVNNVMSRRVYSPMDDLNVSOFLPELSEDSGFKDLNFANYSNNTNRRPSFTF 851
| : : : : :
Db 1347 -----PTYDS-----SEKPEDOKLDVDPVSVKSGTR----- 1374
QY 852 LSTKAVLSINDNDPNVPEPEPKSYAEIRNARRLSANKAPNOAPPLPQROPPSSTRSNS 911
| : : : : :
Db 1375 ---KSLRASLNNEVPAIQ-PEP-----VINTVSSSENSTSKME---DVYTEPNDTK--- 1418
QY 912 NKRVSRFRVPTFE---IRRTSSALAPCDMYNDIFDDFGAGSKPTIKAGMKTLPSMDK 967
| : : : : :
Db 1419 NKRHTRNKSPQIKGSPPTIKHFESIESEAIKGSVDVSKAESSIRPTRQORSY---SKAKS 1475
QY 968 DVKRLNKAAGVTODEYINAKLVD-----QPKKNSIYTDPEDRYEELQOTASIHN 1018
| : : : : :
Db 1476 EFKALADSKSENVSDPSISTYTQDLGQAKYDSEKPSGVDQNLNPKYDSCNE-QDACKDSN 1534
QY 1019 ATIDSSITGRPDSISTDMLPYLSDCLKRPPTALLSADRLFMEOEVPRLRSNSVLVHPGAG 1078
| : : : : :
Db 1535 QAKD-----LPSKRRKROSQSTSKIYPPVSOQPKPEKLHEQNRDESVA----- 1579
QY 1079 AATNSSMLPEPPELINSFARNVSNNSDVAISGNASTISFNQLDMNFDDQATIGQIOE 1138
| : : : : :
Db 1580 -----PKQOLDPPTTSKNSRKGSRSIATNTPTMLTASE---DPEAVEGDSV-E 1627
QY 1139 QPASKSANTVRCDDGLASAPETPR-TPTKKESISSKPAKLSASAPRK-----SPKITG 1191
| : : : : :
Db 1628 VPIKRRKASTRRG-----SVPRDTPISSYVNASDESAVLESV--PKGGRSNISSVPAD 1677
QY 1192 SPVRVIKKNKSAGIEPIPKATHKPK 1217
| : : : : :
Db 1678 TTIPKSSKSDSLNAIEIIPICEPR 1703

Search completed: March 17, 2003, 12:24:42
Job time : 171.627 secs